

Query Match: 70.2% Score 1356, DB 1, Length 430;
 Host Local Similarity: 64.0%; Pred. No. 1,66-109;
 Matches: 275; Conservative: 30; Mismatches: 71; Indels: 54; Gaps: 1;
 DB DISULFID 116 412 BY SIMILARITY.
 DB DISULFID 165 208 BY SIMILARITY.
 FT VARIANT 102 102 G -> E (IN STRAIN N20).
 SQ SEQUENCE 430 AA: 475RS MW: 2PEP1427644404S CWC64.

Query Match: 70.2% Score 1356, DB 1, Length 430;
 Host Local Similarity: 64.0%; Pred. No. 1,66-109;
 Matches: 275; Conservative: 30; Mismatches: 71; Indels: 54; Gaps: 1;
 DB DISULFID 116 412 BY SIMILARITY.
 DB DISULFID 165 208 BY SIMILARITY.
 FT VARIANT 102 102 G -> E (IN STRAIN N20).
 SQ SEQUENCE 430 AA: 475RS MW: 2PEP1427644404S CWC64.

Query Match: 70.2% Score 1356, DB 1, Length 430;
 Host Local Similarity: 64.0%; Pred. No. 1,66-109;
 Matches: 275; Conservative: 30; Mismatches: 71; Indels: 54; Gaps: 1;
 DB DISULFID 116 412 BY SIMILARITY.
 DB DISULFID 165 208 BY SIMILARITY.
 FT VARIANT 102 102 G -> E (IN STRAIN N20).
 SQ SEQUENCE 430 AA: 475RS MW: 2PEP1427644404S CWC64.

Query Match: 70.2% Score 1356, DB 1, Length 430;
 Host Local Similarity: 64.0%; Pred. No. 1,66-109;
 Matches: 275; Conservative: 30; Mismatches: 71; Indels: 54; Gaps: 1;
 DB DISULFID 116 412 BY SIMILARITY.
 DB DISULFID 165 208 BY SIMILARITY.
 FT VARIANT 102 102 G -> E (IN STRAIN N20).
 SQ SEQUENCE 430 AA: 475RS MW: 2PEP1427644404S CWC64.

or send an email to license@slsib.ch.

Query Match: 68.7% Score 1327.5; DB 1, Length 427;
 Host Local Similarity: 64.0%; Pred. No. 1,66-109;
 Matches: 274; Conservative: 34; Mismatches: 53; Indels: 57; Gaps: 3;
 DB DISULFID 116 412 BY SIMILARITY.
 DB DISULFID 165 208 BY SIMILARITY.
 FT VARIANT 102 102 G -> E (IN STRAIN N20).
 SQ SEQUENCE 427 AA: 474220 MW: 2PEP1427644404S CWC64.

Query Match: 68.7% Score 1327.5; DB 1, Length 427;
 Host Local Similarity: 64.0%; Pred. No. 1,66-109;
 Matches: 274; Conservative: 34; Mismatches: 53; Indels: 57; Gaps: 3;
 DB DISULFID 116 412 BY SIMILARITY.
 DB DISULFID 165 208 BY SIMILARITY.
 FT VARIANT 102 102 G -> E (IN STRAIN N20).
 SQ SEQUENCE 427 AA: 474220 MW: 2PEP1427644404S CWC64.

Query Match: 68.7% Score 1327.5; DB 1, Length 427;
 Host Local Similarity: 64.0%; Pred. No. 1,66-109;
 Matches: 274; Conservative: 34; Mismatches: 53; Indels: 57; Gaps: 3;
 DB DISULFID 116 412 BY SIMILARITY.
 DB DISULFID 165 208 BY SIMILARITY.
 FT VARIANT 102 102 G -> E (IN STRAIN N20).
 SQ SEQUENCE 427 AA: 474220 MW: 2PEP1427644404S CWC64.

FT	DISULFID	155	197	BY SIMILARITY												
FT	CARBOHYD	89	89	N-TERMIN (GLYNAAC) () (POTENTIAL)												
FT	CARBOHYD	160	160	N-TERMIN (GLYNAAC) () (POTENTIAL)												
FT	CARBOHYD	165	165	N-TERMIN (GLYNAAC) () (POTENTIAL)												
FT	CARBOHYD	178	178	N-TERMIN (GLYNAAC) () (POTENTIAL)												
FT	CARBOHYD	511	713	N-TERMIN (GLYNAAC) () (POTENTIAL)												
FT	CARBOHYD	427	227	N-TERMIN (GLYNAAC) () (POTENTIAL)												
FT	CARBOHYD	314	246	N-TERMIN (GLYNAAC) () (POTENTIAL)												
FT	CARBOHYD	359	359	N-TERMIN (GLYNAAC) () (POTENTIAL)												
SC	SEQUENCE	116 AA:	46963 MW	338076 EFACRAAE3 (P0664)												
Query Match																
Query Length: 116; Similarity: 22.6%; Posed: 26; DB: 10; Length: 416;																
Matches: 101; Conservation: 42; Mismatches: 119; Indels: 165; Gaps:																
QY	4	ICAFI	PI	EVGAL	TEMP	PERMO	SHE	YKRL	LDQHL	IMMDEL	STDEL	----	NPS	YVGL	58	
DB	9	VGLL	LFST	----	IP	LDQ	DS	SV	VF	KE	IVSH	----	PI	LS	AM	60
QY	59	KE	SS	Q	AE	TR	VE	Y	Y	H	S	I	Q	SS	Q	114
DB	61	RI	AG	IS	Q	SS	Q	AE	TR	VE	Y	Y	H	S	I	105
QY	115	----	N	CE	VF	KE	IVSH	----	PI	LS	AM	----	KE	IVSH	100	
DB	100	RI	AG	IS	Q	SS	Q	AE	TR	VE	Y	Y	H	S	I	105
QY	131	IK	AF	LE	TK	----	AE	TR	VE	Y	Y	H	S	I	105	
DB	156	VY	PI	VE	Y	Y	H	S	I	Q	SS	Q	AE	TR	VE	125
QY	160	FE	KE	VE	Y	Y	H	S	I	Q	SS	Q	AE	TR	VE	125
DB	226	RI	AG	IS	Q	SS	Q	AE	TR	VE	Y	Y	H	S	I	105
QY	219	AK	VA	LA	SS	Q	AE	TR	VE	Y	Y	H	S	I	105	
DB	272	----	IP	LD	Q	SS	Q	AE	TR	VE	Y	Y	H	S	I	105
QY	279	IS	VE	Y	Y	H	S	I	Q	SS	Q	AE	TR	VE	Y	121
DB	319	----	IP	LD	Q	SS	Q	AE	TR	VE	Y	Y	H	S	I	105
QY	322	Q	SS	Q	AE	TR	VE	Y	Y	H	S	I	Q	SS	Q	114
DB	371	PI	VE	Y	Y	H	S	I	Q	SS	Q	AE	TR	VE	Y	125
RESULT 10																
LIPA-CHLPN STANDARD; PRI: 307 AA.																
AC	Q92774	Q9J018														
DI	30-MAY-2000	(rel. 39, Last sequence update)														
DT	16-OCT-2001	(rel. 40, Last annotation update)														
DE	Hydroxylase (LIP SYN) (lipid synthase)															
GN	LIPA OR CPM832 OR CPM38															
OS	Chlamydia pneumoniae (Chlamydia pneumoniae)															
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia															
QX	NCBI Taxid 33558															
RP	SEQUENCE FROM N.A.															
RC	STRAIN-CMD29															
RY	MEDLINE:9920606; PubMed:10192489															
PA	Kimura S, Miyachi N, Murakami E, Imai T, Hyman P, W.															
RT	Comparative genomes of Chlamydia pneumoniae and C. trachomatis															
RI	Nat. Genet. 21:385-389(1999).															
RP	SEQUENCE FROM N.A.															
RC	STRAIN-AR39															
RY	MEDLINE:20150255; PubMed:1064945															

RESULT	14
10A3_HUMAN	
11: D1A1_HUMAN	STANDARD: PRT: 853 AA.
AA: Q9NSV4:	
D1: 16-147T-2001	(Res: 40, Greated)
D1: 16-147T-2001	(Res: 40, last sequence update)
D1: 16-147T-2001	(Res: 40, last annotation update)
D1E: Diaphanous protein homolog 4 (Claphanous-related formin-3) (DREF3)	

Query Match	4.8%	Score 93	DB 1	Length 857
Best Local Similarity	23.58%	Pred No. 4.3		
Matches	52	Conservative	103	Indels
				40
				66ps
				97
179 MATPVEPFIKAGTPEPCEWVCPPLALGRLMISDMGALGTFACIKAPVALASTGD	229			

[illegible]

• ,

OM protein - protein search, using sw model

Paper no.	September 4, 2009, pp. 47-52	© 2009 IEEE, doi: 10.1109/JSTSP.2009.49.638000ds
	(without alignments)	
	946,951 Million cell updates/sec	

11110:	08-09-841-158-3	01ALYB000001111111VSW	476
Period score:	1932		
Sequence	1 MBHGGATTTTCTTTCATTCM		

Gapop 10.0 , Gapext 0.5

seedbed:	562227	562227	562227
Total number of hits satisfying chosen parameter:	562227	562227	562227

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match: 100%
Listing first 45 summaries	

```
Database : SPTRMHL_19:*
```

```

1:  SP_archaea: *
2:  SP_bacteria: *
3:  SP_fungal: *
4:  SP_human: *
5:  SP_invertebrate: *
6:  SP_mammal: *
7:  SP_muc: *
8:  SP_organella: *
9:  SP_phage: *
10: SP_plant: *
11: SP_podent: *
12: SP_virus: *
13: SP_vertebrate: *
14: SP_unclassified: *
15: SP_rvius: *
16: SP_bacteriophage: *
17: SP_archaeal: *

```

SUMMARIES

Result No.	Score	Query Match	Length	PR	ID	Description
1	1779	92.1	400	4	Q66E04	Q66E04 hemG sapin
2	110	5.7	4513	10	Q5MBE8	Q5MBE8 chlamydom
3	107	5.5	844	10	Q5B170	Q5B170 arabinosid
4	102.5	5.3	595	5	Q9AKK2	Q9AKK2 dirosophila
5	101.5	5.3	846	5	Q9Y1H3	Q9Y1H3 dirosophila
6	101	5.2	478	16	Q5B8X0	Q5B8X0 thiobium 1
7	100	5.2	450	10	Q5B8A1	Q5B8A1 streptomyce
8	99.5	5.2	4747	2	Q9L8B4	Q9L8B4 hemG sapin
9	99.9	5.1	1441	4	Q15074	Q15074 arabinosid
10	98.5	5.1	866	1	Q9C594	Q9C594 arabinosid
11	96	4.4	442	4	Q9A0V4	Q9A0V4 arabinosid
12	93.5	4.8	481	1	Q9C714	Q9C714 arabinosid
13	93.5	4.8	747	1	Q9A8V4	Q9A8V4 arabinosid
14	93.5	4.8	794	16	Q9A848	Q9A848 caulobacter
15	93.5	4.8	406	1	Q9A8V4	Q9A8V4 arabinosid
16	93	4.8	886	5	Q6159	Q6159 dirosophila

17	93	4.8	1175	10	Q9SM03	Q9SM03 arabidopsis
18	91.5	4.7	636	6	Q9GJX6	Q9GJX6 sus serota
19	91.5	4.7	643	6	Q9GJX7	Q9GJX7 sus serota
20	91.5	4.7	667	11	Q9RLD9	Q9RLD9 mus musculus
21	91.5	4.7	1719	5	Q9VMR9	Q9VMR9 drosophila
22	91.5	4.7	2162	5	Q9W8V6	Q9W8V6 drosophila
23	91	4.7	1812	5	Q9G9V7	Q9G9V7 leishmania
24	92.5	4.7	1294	2	Q534Z2	Q534Z2 streptomyces
25	90	4.7	2644	4	Q13545	Q13545 homo sapiens
26	89.5	4.6	392	3	Q5K9P9	Q5K9P9 yarrowia
27	89.5	4.6	644	2	Q9K180	Q9K180 streptococcus
28	89.5	4.6	1294	2	Q956R2	Q956R2 streptomyces
29	89	4.6	428	10	Q9FYQ9	Q9FYQ9 arabidopsis
30	89	4.6	1938	15	Q96809	Q96809 ribonuclease
31	89	4.6	2910	5	Q20456	Q20456 caenorhabditis
32	88.5	4.6	717	11	Q9GXJ4	Q9GXJ4 mus musculus
33	88	4.5	5322	5	Q9VLE5	Q9VLE5 drosophila
34	87.5	4.5	379	2	Q501Y8	Q501Y8 mycobacterium
35	87.5	4.5	392	4	Q523Z2	Q523Z2 homo sapiens
36	87.5	4.5	356	16	Q913M3	Q913M3 pseudomonas
37	87.5	4.5	526	2	Q52405	Q52405 pseudomonas
38	87.5	4.5	703	5	Q619V2	Q619V2 trypanosoma
39	87.5	4.5	1206	16	Q9GXX9	Q9GXX9 mycobacterium
40	87	4.5	550	10	Q82347	Q82347 arabidopsis
41	87	4.5	1417	10	Q956G0	Q956G0 arabidopsis
42	87	4.5	2611	2	Q68487	Q68487 streptomyces
43	87	4.5	4390	5	Q9GYD2	Q9GYD2 drosophila
44	87	4.5	5392	5	Q9GYD2	Q9GYD2 drosophila
45	86.5	4.5	284	14	Q918H0	Q918H0 xenopus laevis

ALLEGMENTS

RESULT	1
G96FH4	
ID	G96FH4
PRELIMINARY:	PRT: 400 AA.
AC	G96FH4;
D7	01-DEC-2001 (TREMBLE). 19, Created)
D7	01-DEC-2001 (TREMBLE). 19, Last sequence update)
D7	01-DEC-2001 (TREMBLE). 19, Last annotation update)
DE	UNKN-WN (PROTEIN POP MCC:17407).
OS	Homo sapiens (human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eumetaria; Primates; Catarrhini; Hominoidea; Homo.
OX	NBI_Laxid=9606;
RN	
RP	SEQUENCE FROM N.A.
RS	1836; GRAFT, AND ANK-LAILE Grafting-Establish MA with 16,19g LOSS;
RA	Strausberg RJ;
F1	1-Jan-1993 (J02717) - The EMBL database, 1993 databases.
DR	EMBL: BC011249; NIH11249.1;
SO	SEQUENCE 400 AA; 44421 MW; 36C59B8C61BC15D CRC64;

Genetic Match	92.1%	Score 1779	DB 4	Length 400
Best Local Similarity	68.8%	Prod. No. 2.2e-16		
Matches 355	Conservative	7	Mismatches 14	Indels 24
				Cuts 2

QY	1	MRHGAFLPLGVIAATTPMPT	PMRSHI	VEKIQHLL	PMWBL	SLPHI	NPST	YVGL	60
Db	1	MRHGAFLPLGVIAATTPMPT	ELPMDSH	VEKIQHLL	PMWBL	SLPHI	NPST	YVGL	60
QY	61	SSLGAEFLPLVYHFL	SYG	ELQSA	NSG	SG	QSG	PMGL	AYLAL
Db	61	SSLGAEFLPLVYHFL	SYG	ELQSA	NSG	SG	QSG	PMGL	AYLAL
QY	121	DRF							
Db	121	DRF							
QY	177	SSNRPLPSSGATKAL	EVET	EL	EAQ	VE	DRH	SNYS	PLAL
Db	177	SSNRPLPSSGATKAL	EVET	EL	EAQ	VE	DRH	SNYS	PLAL

Query Match	5.38	Score 102.57	DB 5	Length 595
Best local similarity	21.39	Prod. No. 212		
Matches	76	Conservative	43	Mismatches 138; Indels 99; Gaps 13
QY	99	KP	SMGLATYLLALPAREYKIKRQVYS-----	129
DB	243	KPIIDFELIAYVALISICIKRQVHNSHVAITGHEFAQTFPAITISASSANHP	302	
QY	139	QVETETETK	FATLFAAKAATFQFESHTNDESEETMAITVEETIKAO	183
DB	403	KKQIPIQITDIASQVNSQVATAMVIAVETVTH-----PHRIQIIVRAPIKATIQ	358	
QY	184	TPHPTQNNVSTPAIQETMSPMCAGETIATKAPVALIASIQAFQNALMISQIP	243	
DB	459	DQKRSQSRKSTALAMQATQIQEPDAPKHWNTASVPIISQPAKQNSQEPPIQKQDEP	418	
QY	244	VINKETVDDLP-----DQ-----LAPVYLEAFETIQ	274	
DB	419	ITNGSLTADITIALCKKGLAAYPAQDQVIVESQITENSEETIAPVPLESSAFSSQA	478	
QY	275	QVETISVTVQV-----ISGLDPRQSTISVLACSVETKKAQETAGVETQAS	324	
DB	479	ENIAYILAVCNVILSLGKAVETETET-----FATLFAAKMDETFEAKNS	530	
QY	425	LSQVYLSNVCVQADEP-----EPWQIP-----QVNTIITQ-----IAYVPIVQK	367	
DB	531	PNQIVHITLYGKKEPPQVHWLTPLEPQNPNTPEISGLIAYVIVETVETVSKP	586	
RESULT	5			
QVYI13				
ID	QVYI13	PEPILIMINAPY	PRF:	836 AA.
AC	QVYI13			
CT	01-MAY-2000 (TEMPREL 13, Created)			
CI	01-MAY-2000 (TEMPREL 13, Last sequence update)			
CF	01-JUN-2001 (TEMPREL 17, Last annotation update)			
DE	CG9342 PROTEIN.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Ecdyzoa, Hexapoda, Insecta, Hexapoda, Insecta			
NC	Phylogeny, Neoflyca, Endopterygota, Diptera, Brachytera, Muscomorpha			
CC	NCBI, TAIR2 7227			
RR	11			
KP	SEQUENCE FROM N.A.			
EC	SPRAIN INTEREST			
RA	MEDLINE 20176006; PubMed 10731132			
FA	Adams W.D., Conboy S.E., Holt P.A., Jones C.A., O'Keefe J.D.,			
FA	Amantides P.G., Baker S.E., Li F.W., Backus R.A., Galle P.F.,			
FA	George B.A., Lewis E.D., Richards S., Ashburner M., Ingham S.N.,			
FA	Gitter J.C., Wirtzen J.F., Vachtel M.D., Chang G., Chen L.Y.,			
FA	Pringle A.C., Rogers Y.H.C., Ricci P.C., O'Keefe M., Pfeiffer B.D.,			
FA	Man K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Melios G.L.G.,			
FA	Abell J.P., Atiyanti A., An H., Andrews Flannock C., Baldwin D.,			
FA	Balcer E.M., Bate A., Kawachi J., Kugler-Fogel L., Garske E.M.,			
FA	Bickson K.Y., Benos E.V., Betman R.L., Blumfeld D., Bolstad S.,			
FA	Borkov D., Borkan M.R., Bonck J., Brakstein P., Bruller P.,			

BA Burtis K.C., Busam D.A., Butler H., Cadion F., Center A., Chandra L.,
 BA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
 BA de Pablos R., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 BA Dodson K., Doup L.R., Downes M., Dugan-Kocha S., Duvkov B.C., Dunn P.,
 BA Durbin K.J., Evangelista C.C., Ferraz G., Ferreira S., Fleischmann W.,
 BA Foster C., Gabriellian A.B., Gary N.S., Gelbart W.M., Glasser K.,
 BA Goff A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
 BA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Hawk J.,
 BA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Itham G.,
 BA Jullis M., Kalish F., Karpen G.H., Ke Z., Kronison J.A., Ketchum K.A.,
 BA Knebel R.E., Knebel C., Kratz G., Kraybill S., Kuf D., Lal Z.,
 BA Laszlo P., Lee Y., Leventy A.A., Li J., Li Z., Liang Y., Lin X.,
 BA Lu X., Maitel B., McIntosh T.G., Meleod M.P., Meperson D.,
 BA Merrett G., Miskin R.V., Mobley C., Morris G., Mostoslav A.,
 BA Merz S.M., Mey M., Murphy B., Murphy L., Nagy D., Nelson D.L.,
 BA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacheb J.M.,
 BA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
 BA Poirier K., Robinson K., Saunders R.D., Schaefer F., Shen H.,
 BA Shup R.C., Shon-Klamis T., Simpson M., Skupski M.P., Smith T.,
 BA Spier F., Spadling A.C., Stapleton M., Strong P., Sun F.,
 BA Sztachalski R., Tabor L., Turner E., Winkler A.B., Wang X.,
 BA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weinstein J.,
 BA Williams S.M., Woodage T., Worley K.C., Wu P., Yang S., Yao Q.A.,
 BA Ye J., Yeh K.-F., Yeh J.S., Yeh M., Yeh M., Yeh S., Yeh X.,
 BA Yeh X.H., Yeh F.R., Yeh W., Yeh X., Yeh X., Yeh X.,
 BA Gibbs R.A., Myers E.W., Rubin G.M., Venturi J.C.,
 BA "The genome sequence of *Drosophila melanogaster*,"
 BA Science 287:2185-2195(2000).
 BA EMBL: A6004668; AAF59456.1;--
 BA Flybase: FB00032904; CG9442.
 BA InterPro: IPR001005; MYD_JNA_bind.
 BA PROSITE: PS00047; MYR_1; UNKNOWN_1.
 BA Sequence: 846 AA; 93040 MW; B4761372AF4F0D14 GR664;

Query Match: 5.38; Score 101.5; DB 5; Length 836;
 Best Local Similarity: 19.48; Pred. No. 4.3;
 Matches: 96; Conservative: 69; Mismatches: 169; Indels: 163; Gaps: 17;

DB 8 LELAVIALFEMCEIPENDSHVERKQGHILPMMDI SIHINPSIYATPISTQAG 66
 12 LIIATLPIGLE-----DEKALLAPNSGQTEKQWVILGEL-----GRSSSAETV 60
 QY 67 TKEVATLNSKICVQGLTASASSELGAPGPK-----TSMQALYALALANDEPVAG 121
 DB 61 TFEDEKINWSTCEGLLEFISQSKVASCASSTLEIEGQYIELVQGPQKVA 129
 QY 122 H-----KIDRIYSLQMLLEENKALHILAMAGIATLKA----- 159
 DB 121 HTSPQGLNIPFETANLQPLIPASOHELEVSSECPVSVNVSSTVEKTRPQSLM 179
 QY 159 --RSNFGKQKPTMAITVPEF-----LKAQTFEHEFNVSSTLALQFLMTS-- 205
 DB 180 DLKAVNHIHALVQVQAGETVEYELSDILLHNSGHEHPIAAFTVSVFVSSIT 239
 QY 269 -----EPKASGTA----- 322
 DB 240 LQHVSGSEHVKQIGQSTPKAIGSTFVNVPEFTSVWQMTAFEGDTETQKAST 299
 QY 277 LLAELQKAFQNALMISQLP-----VLNKKTYDITITQGLAFVMEFPAETFOGRT 279
 DB 300 FLOSNQVKSALAVKLLPIARITGQGFED-----LTERAVIPLQVLDL 348
 QY 280 -----SVTLQVSLTPPEQSLV-----LANSIVEVA 368
 DB 349 GAVQTPAHNATFGLYKESSETSEGLLEKYLQSLAVATHDRKIVPHGLLEPQBSI 408
 QY 369 KKAHFGATFTFTASISQYITSVKAKAGEFEFVLLPEFPIPLGSLAIYEP----- 363
 DB 409 KKKHLEPSVIGVAVIT-----PQSHIVPEPIIKFVPSVITPGLTSKEPTLYIP 459
 QY 464 -----KIKETLERL 373

DB 460 ALQNDQPAIEAL 474
 RESULT 6
 ID Q98GYO PRELIMINARY: PRT: 478 AA.
 AC Q98GYO:
 DT 01-OCT-2001 (11FEM01 18, created)
 DT 01-OCT-2001 (11FEM01 18, last sequence update)
 DT 01-OCT-2001 (11FEM01 18, last annotation update)
 DE MLE325 PROTEIN.
 CF MLE325.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAF304099;
 RC MFLINP=21082940; PubMed=11214968;
 RC Karche I., Kishida Y., Sakaguchi Y., Sakaguchi Y., Yano T., Yasumoto S.,
 RA Kishida Y., Kiyokawa G., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakajima R., Shimizu S., Sugimoto M.,
 RA Takashima G., Yabuta M., Yabuta S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti."
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003001; BAB50086.1;--
 DR InterPro: IPR001958; TCR_LYOA.
 DR InterPro: IPR001411; TCR_TOYB.
 DR PRINTS: PRO1035; TCRTOYB.
 DR PRINTS: PRO1036; TCRTOYB.
 KW Complete proteome.
 SO SEQUENCE 478 AA; 49556 MW; 44725062780894 GR664;

Query Match: 5.28; Score 101; DB 16; Length 478;
 Best Local Similarity: 24.18; Pred. No. 2.2;
 Matches: 106; Conservative: 49; Mismatches: 165; Indels: 140; Gaps: 24;

QY 4 LGAPLFLTGVIGALFEMCEIPENDSHVERKQGHILPMMDI SIHINPSIYATPISTQAG 61
 DB 20 LAFPEHSLQGLHVAHETVAVAVIMPAVWVHAGVQVWV-----SLATV-----ELG 67
 QY 62 STACIETKEDYELHSLQVQGLTASASSELGAPGPK-----TSMQALYALALANDEPVAG 121
 DB 68 ALVAGAAAGHVSVALKIMVVAALLAAGALICATISM-----QELTAG-----KLIEG 119
 QY 122 HPGPIVSGCKWPLEPK-----KAI-----DIAAMAGIATLKA-----KSNF 160
 DB 120 LGGALVSLA-----EVSVERLEFSAIPQLEQIMSALWGVAFSSQDIALALMPEISWMAF 177
 QY 161 NPGPPIITMAPIVPEPFIACFPGHEHNVSTPAIGFTMSMPEAFELTAKAR 220
 DB 179 GVTGLGATIMLAS-----PIVIMPP-----AKRISTAG-----PIHPEFAAL-----ACTAVA 224
 QY 221 VAIASLQVAPQAMTSQLIVLNERTYDLEFPEVAPVVMPIA----- 268
 DB 225 VLIAS-----ASVDIALKSSLLIVLAGIALFFYIHALPKSRILPABLSWRIWAG 281
 QY 269 -----AFTIP QIGRII-----SVTLQVSL-----PPVQS 295
 DB 282 MIVAAEVAIADSTGVQPLILSLHETVITGVYIHAESTAWSTSLVAVNAFQPKR 341
 QY 296 ISVAGTVEVDLKAHLEGGFTYFTQ-----ANISQ-----PYLSVMKKA 348
 DB 342 LIIITSA-----LMIAGINGAVATIPIDISPLLICALLQGGRIAMPIIIVIVASA 396
 QY 399 GEPEFWLLPQNPPIPLQGI 358
 DB 397 PDEE-----QTIASAAPVIMORI 414


```

RESULT 7
Q9CAH1 PRELIMINARY: PRI: 450 AA.
ID Q9CAH1
AC Q9CAH1
DI 01-JUN-2001 (TREMBL: 17, created)
DI 01-JUN-2001 (TREMBL: 17, last sequence update)
DI 01-OCT-2001 (TREMBL: 18, last annotation update)
DE HYDROLIC 59.9 kDa protein.
GN F2822.27.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicot; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI:Taxid:3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:CV, COLUMBIA;
RC MEDLINE:21016719; PubMed:1110712;
RA Bhoolajis A., Porter J.P., Palm G.L., Federoff N.A., Earl S.,
RA White O., Alonso J., Altieri H., Arango P., Bostwick S.T.,
RA Biedler E., Chan A., Chao Y., Chen H., Chou K., Chin C.W.,
RA Chung M.Y., Cunniff C., Gentry A.B., Gentry A.B., Gentry T.B., Gentry K.,
RA Dunn P., Fagan P., Faldutsky V., Fong C.B., Fong B., Fujii C.Y.,
RA Gill I.P., Goldsmith A.P., Haas P., Hanson N.F., Harber B., Hatcher L.,
RA Horner T.L., Jankins J., Johnson-Hopson C., Khan S., Khaykin F.,
RA Kim C.T., Koo H.I., Kremenetskaia I., Kurle D.H., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lee C.A., Li J.H., Li Y.F.,
RA Lin X., Liu S.Y., Liu Y.A., Lures J.S., Maiti P., Matelli A.,
RA Miltner J., Miranda M., Nguyen M., Norman W.C., Osborne B.L.,
RA Pat G., Peterson J., Pham P.Y., Pizarro M., Pomeroy T., Pomeroy D.,
RA Sapiro H., Sapiro S.I., Schwartz J.P., Shier P., Siedlitz A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tofuni M.J., Town C.D.,
RA Uiterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
DR EMBL: AC010926; AAG51840.1;
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002900; Ser_Thr_Pkinase.
DR InterPro: IPR012451; Tyr_Pkinase.
DR Pfam: Pf00069; Pkinase_1;
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00220; S_TKc_1;
DR SMART: SM00219; Tyrc_1;
DR PROSITE: PS00307; PESTIN_KINASE_A; CNRKNRL1;
DR PROSITE: PS00011; PROTEIN_KINASE_DOM_1;
KM ATP-binding: Hypothetical protein; Transferrase;
SU SHODNEP 460 AA 7997 MW 84379.69366 kDa 59.9%.

Query Match 5.28; Score 100; DB 10; Length 450;
Host local similarity: 23.94; Proc No. 25;
Matches 56; Conservative 27; Mismatches 71; Indels 90; Gaps 11;

```

```

DB 379 VIVA 382
RESULT 8
Q9L8H4 PRELIMINARY: PRI: 4247 AA.
ID Q9L8H4
AC Q9L8H4
DI 01-OCT-2000 (TREMBL: 15, created)
DI 01-OCT-2000 (TREMBL: 15, last sequence update)
DI 01-DEC-2000 (TREMBL: 19, last annotation update)
DE ACTINOMYCIN SYNTHETASE 111.
GN ACTM.
OS Streptomyces chrysomallus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomyces; Streptomyces.
OX NCBI:Taxid:1899;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN ATCC 1523;
RC MEDLINE:2319879; PubMed:1078924;
EX Solonowick E., Pomeroy K., Gammie N., Miller D.;
RA "Characterization and in vitro analysis of a new 66 kDa modular polypeptide
RA synthetase for synthesis of N-methylated argyl peptides."
EL Chem. Biol. 7:287-297(2000).
DR EMBL: AF294101; AF1474.1;
DR HSP: P14687; LMO;
DR InterPro: IPR00873; AMP-bind.
DR InterPro: IPR001342; DOP4.
DR InterPro: IPR003479; Est_Lip_Thioest_acsite.
DR InterPro: IPR01601; Meth-transf.
DR InterPro: IPR031889; Thioestant_attach.
DR InterPro: IPR000051; SAM_bind.
DR Pfam: PF001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 3;
DR Pfam: PF00464; Condensation; 3;
DR Pfam: PF00475; Thioesterase; 1;
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00075; ACT_DOMIN; 4;
DR PROSITE: PS00455; AMP_BINDING; 3;
DR PROSITE: PS00912; PESTINPARTITHEIN; UNKNOWN_3;
KM Thioesterantichine;
SU SEQUENCE 4247 AA 46229 MW 58067.91522 kDa 59.9% CDS64;

Query Match 5.28; Score 95.5; DB 2; Length 4247;
Host local similarity: 22.98; Proc No. 60;
Matches 61; Conservative 37; Mismatches 109; Indels 59; Gaps 13;

```


AC 015074:
 DT 01-JAN-1998 (Tremblérel, 05, Created)
 DT 01-JAN-1998 (Tremblérel, 05, Last sequence update)
 DT 01-OCT-2001 (Tremblérel, 19, Last annotation update)
 DE KIA0368 PROTEIN (FRAGMENT)
 GR KIA0368
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catartini; Homiidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE BRAIN;
 KA Medline 9744984; PubMed 920541;
 KA Nagase T., Ishikawa K., Nakajima O., Ohira M., Soki N., Miyajima N.,
 KA Tanabe A., Kodani H., Nomura N., Ohara O.,
 KA Prediction of the coding sequences of unidentified human genes. VII.
 KA The complete sequences of 100 new cDNA clones from brain which can
 KA code for large proteins in vitro.
 RL DNA Res. 4:141-150(1997).
 RL EMBL: AB003464; MAA20824.1;
 DR InterPro: IPR004015; HLA_Myc.
 DR PROSITE: PS00092; HELIX_LOOP_HELIX; CRRKRN_1.
 FT NON_TER
 SQ SEQUENCE 1441 AA; 159477 MW; 55925574386164 CRC64;

Query Match 5.1%; Score 99; DR 4; Length 1441;
 Best local Similarity 20.0%; Pred. No. 15;
 Matches 84; Conservative 66; Mismatches 143; Indels 126; Gaps 22;
 2 RHAAFLFLAVGALTEMETPEMUSHLVEK-----GQHLHWDRK-----SLEHIN 51
 DB 845 RTIAALLPILDKGMSTVEFKALSTNTVAKISKASGMLKPIAPKLIPALLLSYLE 904
 QY 52 PSL--YVGLRISLQAGTKEDLYHSKIG-----YQGL-----LGSAPSE----- 91
 DB 905 PVLNLTSLKAPLPKAAUSARLSAASSPMKEETIMLAVLYSVGLVPELCELEIR 964
 QY 32 -----DKKVGQ-----KPMGLAVYLA-----LPANPEVPGH-- 122
 DB 955 SNGDGLPDEKASVAVSTLPFGDLYSTPMASLSTLHNSVLYGSAFAMGHPV 1024
 QY 124 -----KQGLVSGLR-WHLLEKVALDIAAAGLAFGLKPSNPQ-----PPQPTM 170
 DB 1025 RLSHSSTFKLQKLNKMWKEELPLKISC-----ALLHAGKSPDLKMKHKEVPL 1080
 QY 171 ALITVEP-----ELIKATPEGHCNYSITPLALQELMTSMRGAELEGT 214
 DB 1081 APLGMRILADEKSEKERNIMTEWQENVP-GSPGGL--KLYQELITITQKALQSOS 1136
 QY 215 ACLKARVAL-IASLDGAFUNA-----LMTSQLPVLNHRV-----IDLTP 256
 DB 1137 WKRRAGATAMASIT--AKQTSIVPRYIGMLITALLQGLAGRIWAGREILKALACVVT 1193
 QY 257 DGLAPVMLPAAETIPQGETISVTLVNLSTLPYPSI--SVIAGSTEIVVKKARF 313
 DB 1194 ACSAE--LEKSVNPGNSINELQAVLKPS-----KENVKRYVAISCAADILKAKE 1244

RESULT 1)
 Q96594
 ID Q96594 PRELIMINARY: PRT; 856 AA.
 AC Q96594:
 DT 01-JUN-2001 (Tremblérel, 17, Created)
 DT 01-JUN-2001 (Tremblérel, 17, Last sequence update)
 DT 01-OCT-2001 (Tremblérel, 19, Last annotation update)
 DE MCA-TRANSCRIPTIONAL ACTIVITY-LIKE PROTEIN.
 GR AF662193.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicotyledons;
 OC Rosidae; Malvales; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID:3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 KA Bevan M., Murphy G., Ridley P., Hudson S., Bamford L., Mewes H.W.,
 KA Rudi S., Lemcke K., Mayer K.F.X.,
 KA Submitted (Mar 2001) to the EMBL/Genbank/Trna databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 KA Arabidopsis sequencing project.
 KA Submitted (Mar 2001) to the EMBL/Genbank/Trna databases.
 RL EMBL: AL589884; GAC4446.1;
 DR InterPro: IPR001757; E1-E2_Alipase.
 DR InterPro: IPR001454; Hydrolase.
 DR Pfam: PF00122; E1-E2_Alipase; 1.
 DR Pfam: PF00702; Hydrolase; 1.
 DR PRINTS: PR00119; CATIPASE.
 DR PROSITE: PS00154; ATPIASE_E1_E2; CRRKRN_1.
 SQ SEQUENCE 856 AA; 90843 MW; 113127345800646 CRC64;

Query Match 5.1%; Score 98.5; DR 10; Length 856;
 Best local Similarity 21.0%; Pred. No. 8.1;
 Matches 75; Conservative 42; Mismatches 145; Indels 74; Gaps 18;
 28 SILVERLDGHLF--WKEPLSLHNPSIVGLPESLVAETKEDYTHSTFLSYGCL 84
 DB 168 SHLHSLGHLHAGGLW-DLHNSYKGLAVGALL-----GREGELFLDKIK-- 214
 QY 85 LGSATSLWQDQGRKSGQALVILALPANVIVPGHVPVYSGIKWLEHKKALPT 144
 DB 215 --AFGRSPNNNSIVGLSMAAPSS-----LISVPEKMAAPSE 206
 QY 145 AWAAGLAFGLKPSNPQPPGLIMAFVPEPTKQGLPEHGVNYSITPAQPM 204
 DB 257 PVLNLTLPVLLGS-----LEERAKLGASTNNEL-----SLSTQSL VIT 208
 QY 205 SPWGAFFCTAC KAPVALLASLGLVAFVNAIMISQLPVLNHRVITLITP--LPLAP 261
 DB 299 SDNNTPVDSVLSDSICINVSVD-----LRVDSLVLNGET-----FVMSVLAG 347
 QY 262 PVLNLPVPA--APLPDQ--PGLISVIVGMSLDPYFGSTVAVASVGVLEFAHLEAG 316
 DB 348 ESWVTESMITGSEIVVPEPFGVAVGATINWPTIPFASSTGKSTSTKLVAVNEDAG 407
 QY 317 FIVETGA--SLSGVLTSGVMCKAAETPEFWLDEPNT--LIGLAVVYKKEETLE 371
 DB 408 NAAPVGRIMADLAGPVVYTIMSLMTAFWYVYSHIFPVLLNDIAG PDAIAL 464
 QY 372 KL 373
 DB 465 SL 466

RESULT 1)
 Q9BDV4
 ID Q9BDV4 PRELIMINARY: PRT; 942 AA.
 AC Q9BDV4:
 DT 01-JUN-2001 (Tremblérel, 17, Created)
 DT 01-JUN-2001 (Tremblérel, 17, Last sequence update)
 DT 01-OCT-2001 (Tremblérel, 18, Last annotation update)
 DE BRCA1 (FRAGMENT).
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID:9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 KA Medline 21082981; PubMed 11214418;
 KA Amodeo O., Scally M., Bogard C., Kao L., Leary K., Atkins K.,
 KA Amodeo H., Stanhope M., de Jong W., Springer M.,
 KA Parallel adaptive radiations in two major clades of placental
 KA mammals.
 RL Nature 409:610-614(2001).

KP STROPHNOCEROM N.A.
 RX STRAIN-CV, COLUMBIA
 RX MEDLINE-21016719; PubMed-11110712;
 RX Theologis A., Eckert J.R., Palm C.J., Federspiel N.A., Kunt S.,
 KA White O., Alonso J., Altieri H., Arango J., Bowman C., Brooks S.Y.,
 KA Brähler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin G.W.,
 KA Chung M., Chen L., Cheng A.R., Chinn A.F., Christy T.H., East E.,
 KA Dunn P., Egan P., Feldhahn T.V., Feng T.D., Ford R., Fujita Y.,
 KA Gill J.E., Goldsmith A.D., Haas R., Hansen M., Hughes B., Hutzler L.,
 KA Horvath T., Inoué T., Janda K.R., Jansky J., Kim E., Kishimoto T.,
 KA Knecht T., Koo H.L., Kromaschik J., Kunkel K., Kunkel A., Lander
 KA Landin-Hogoper S., Lee A., Lee J.M., Lenz G.A., Liu J.H., Li Y.P.,
 KA Lin X., Liu S.X., Liu P.A., Lyons J.S., Katali R., Marshall A.,
 KA Miltner T., Miranda M., Nguyen M., Nguyen W., Ostergaard E.,
 KA Pail G., Petersen T., Pham V.K., Pitzer M., Poggio T., Pong P.,
 KA Spagno H., Storz S.L., Schwartz J.R., Shinn P., Solowick A.M.,
 KA Sun H., Tallon L.J., Tamminga G., Tortum M.J., Town G.D.,
 KA Vetter J.T., Vu A.H., Vyas G., Vyas G., Vyas G., Vyas G., Vyas G.,
 KA Wu D., Yu G., Fraser G.M., Venturi J.C., Davis R.W.,
 KA "sequence and analysis of chromosome 1 of the plant Arabidopsis
 KT thaliana."
 KA Nucleic Acids Res (2000)
 CV - SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 RX EMBL: AC074025; AAC51161.1; -
 RX InspecPro: Ipp001128; CV: 1450.

RX STRAIN 0266/952. PubMed-11139578;
 KX MEDLINE 21179132.
 RA Van G., Luo H., Lee J.-D., Abe J., Berk B.C.:
 ET "Molecular Cloning of Mouse EEF2/SME1 Effector Variants and
 ET Characterization of EEF2 Functional Domains." *J.
 KJ Biol. Chem.* 276:10487-10493(2001).
 C1 J. STEVENS, B. HARRIS, J. L. SELL AND THE FAMILY OF EEF2/EFEN KINASES.
 DR EMBL ALMANAC: AAD019355.1. *
 DR HSNP: Q16539; IMPC.
 DR NCBI MIM:141447; RefSeq.
 DR UniProt: P00009; EC: 2-kinase.
 DR UniProt: P000327; KMF:kin.
 DR UniProt: P0002965; Pfam:extensin.
 DR UniProt: P0003716; PRA:pol omega.
 L5 UniProt: P152293; Ref:the_kinase.
 DR Pfam: PF00069; Pfamase: 1.
 DR PRINTS: PR01217; PROSITEEXTIN.
 DR SMART: SM00220; S:kin: 1.
 DR UniProt: P001451; MAF: 000749.1.
 DR UniProt: P05011; PROTEIN KINASE ST: 1.
 DR UniProt: P000108; PROTEIN KINASE ST: 1.
 DR AIR: Binds to kinase, serine/threonine tyrosine kinase, transferase.


```

Query Match: 4.88; Score 93.5; DB 11; Length 806;
Host Local Similarity 19.28; Pred. No. 20;
Matches 83; conservative 51; Mismatches 128; Indels 171; Gaps 19;

QY 22 FTEDMSHLVEKLEGGHLEMMOR.SLEHNPSTYVCHLSLQAGLEKHGYSLSKISQY 81
DB 138 DIMSDHJHJHSQ-----PILTEHYRYLYQILK-----GLK---YMSAOVHR 181
QY 82 GQIIGSASFEDGGGQKPSMQQALYLIALPANCPEVPCQK-----LELVS 129
DB 182 FY-----KERR-----LIVRECHKLEFGMAKLESTFAHREYET 219
QY 130 Q-----LKNF-----LEKKKALDTAAGLAFTLKPSNENPCP-----QRLIMAIR 173
DB 220 FYVAFQWPAPELMSLEHVTQALPMSVQYIIEHMAPEGLTQRYVHIGLIMVGLG 279
QY 174 TVREHLLKAGTFECHGNAVSTP-----LACGIMETIMCALIGTA 215
DB 280 TPSIAVIGQAVAFVAVATQSLPRQGPVMEYVPCAHGALSLSPMLPEFSAPISAA 339
QY 216 GL-----KAVVALLASLQD-----GAFQNA 235
DB 340 AALRRPEIATYRRHRRHRRFAPPEVAFQEFALIEETLEFATVAFIPFVAPPEGLPQOI 399
QY 236 LMSGLAPVNHKTYTHLP-----DGL--APRVMLHTAETTVQFQHLISVTLQ-- 284
DB 400 KFOPLQPVVSEPVCPHVFMSPPSGDCAMESPPALPPGSDPAPPT--VDITLQPA 456
QY 295 VSLIPEYQSLSVATVEILFAHHLQPEFYAGASLSQGLISWGLPAGGRK 342
DB 457 PPSHLATPKEE-----GAINSLIAALAAALNLSGR----- 490
QY 443 FVGLLRDINTPL 354
DB 491 ----LRKPSAPL 499

```

Search completed: September 4, 2002, 11:48:26
 Job time: 7624 sec

.

.

.

lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer.

PS claim 11: Page 954-955; 1425pp; English.

XX Polynucleotide sequences AAF12982 - AAF1424 encode human lung cancer
 CC associated proteins represented in AAF58106 - AAF58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective, cytoprotective, cardioprotective,
 CC immunomodulatory, muscular active, general, vitelline, gastrointestinal
 CC activity, nephroprotective, anti-infective, gynecological, or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF1425 - AAF1433 and
 CC peptide AAF58549 are used in the course of the invention for the
 CC identification and characterization of the polynucleotide and protein
 CC sequences.

XX Sequence 268 AA;

SV

Query Match

1st Local Similarity 41.6%; Score 803.5; 108 21; Length 268;

Match 1st Conservative 0; Mismatches 4; Indels 51; Gaps 1;

XX

YY 1 MRHGAFLFLIGVAGLTEMCEIPEMDSHLYEKIGQHLPMRSLSEHLNPSIYGRU 60

DB 39 mnhgaflllllqvaltemceipemshlyeklqhlplpmdrlstehlnpsiygrll 98

YY 61 SSLDAGTKPKYVYHSTKYGQGLTASAPSEDCGPGGKRSWGALLYLALPANCEFYR 120

DB 99 ssldagtkedylhstklqyqcllqsalreddqgkpsmgqlasxllalrancelvx 158

YY 121 GHRKPIVLSQIKWLEDEKPAI 142

DB 159 ghrkplvlsqikwledekdshlqphlssygytlllslhphkrlvdsydkll 218

YY 143 -----DTAAMAGIAFTCLKRSNPNQKR 165

DB 219 yavepfbqphhsydamagialftclkrslpqr 254

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

XX Dumas Milne Edwards J, Imperant A, Giordano J;
 DB WPI: 2000-500481/45.
 DB N-PSDB: AAC00291.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 CC obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures
 CC Claim 13: SEQ ID 4366; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or poly(A)⁺ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 54 AA;

SV

Query Match

1st Local Similarity 14.8%; Score 285; 108 21; Length 54;

Match 1st Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

YY 1 MRHGAFLFLIGVAGLTEMCEIPEMDSHLYEKIGQHLPMRSLSEHLNPSI 54

DB 1 mnhgaflllllqvaltemceipemshlyeklqhlplpmdrlstehlnpsl 54

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

DB

YY

119	02-MAH1993.
XX	
FE	06-JUN-1991; 91JP-0134788.
XX	
FE	06-JUN-1991; 91JP-0134788.
XX	
PA	(KANP) KANEKA CORP.
XX	
DE	WEI; 1993-11188/14.
XX	
PT	DNA sequence coding human intrinsic factor - is used for cloning
PT	human intrinsic factor DNA using the new DNA as a probe
XX	
PS	Disclosure; Page 9-10; 11pp; Japanese.
XX	
XX	This sequence represents rat intrinsic factor. This sequence was
XX	isolated using probes based on the human intrinsic factor cDNA.
XX	
SC	Sequence 421 AA;

Query Match 13.0%; Score 251.5; DB 14; Length 421;
 Best Local Similarity 23.48; Pred. No. 5.2e-18;
 Matches 104; Conservative 70; Mismatches 146; Indels 125; Gaps 17

```

0Y 6 ELLEVALGAT-----EMCEIPENOSHVERKOGHILMORPST-----ELIND 52
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 11 YILINOLKASASTIATASOPFADAT-----POTOGATIMASSTADIP 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 5 SIVVALASTIGATPEFVYHSHKOGYOGATLOSTASSTIGANOSKESMOMANIVAT 112
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 60 SILLIMILAS-----LYNLEAKILLYIMASSTADITOGALIMAL 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 11 KANPEVCHOKHOKHVSOLKAW-----133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 104 TSETTAPKOSSTIPKIMESPESTIGASSTIYOGALILLOKSEVALPILATTA 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 134 FLEDERALDIPAMAGIATLO-----KSNENOSPORITMAITPEPEFIRKAO 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 16 KIMOSPESTIGATILALIMOTPEFISAGYOGYEDITIGALILLYANISIT 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 184 TPECHONONYSTALOGILMISPEKHAISTATAPAVNASTHOCAPANMISTIT 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 216 KOPALITIGYOGATASSTIPKOLKOSKOSKIMILYOGYOGATIMASITIP 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 244 VIMHEKYITLIPYILAPVIMPEAFETIPOTGE-----LISVT-----LOVIS 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 278 SIKAGKYILK-----PYTCOPDOPPEITIGYPTFUSISITILYIMINILY 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 286 LIPYKOSTIVASTIVIMPRANHEONITIVIGASTONILYV-----MKKACPEHEW 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 331 LITVITPEVSKASVILYALOGAFRIMKIKOLUMESKOWIVSSIMILANIKKLYW 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 345 GILPONTIPILANIPKKEPIT 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0b 391 ELSKIPKOSQAYVYFIRYOGIT 414
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT	6
AAV*5475	
ID	AAV*5475 Standard; Protein; 319 AA.
XX	
AC	AAV*5475;
XX	
D1	13-SEP-1999 (first entry)
XX	
DE	Chlamydia pneumoniae cellular envelope protein.
XX	
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW	vaccines; neutralising epitope.
XX	
OS	Chlamydia pneumoniae.

XX
EN WC9927105-A2.
XX
FD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-1R01890.
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97ER-0014673.
XX
FA (GHST) GENSET.
XX
PI Griffiths R;
DP WPI : 1999-357842/10.
XX

Genome sequence of Chlamydia pneumoniae

PAGE 1239 1240; Disclosure: 1012PP; English.

AAW34584-Y35879 represent the proteins encoded by all the open reading frames in the complete genome (see AAW34584) of Chlamydia pneumoniae. C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AAW34584-Y35879) can be used in immunologic compositions as vaccines. Vaccines containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector affects the expression of a neighbouring epitope of C. pneumoniae.

Sequence 119 AA;

Every Match 5.48; Score 104.5; TB 20; Length 819;
 Best Local Similarity 22.48; Pred. No. 0.031;
 Matches 72; Conservative 47; Mistakes 115; Cops 18;

07	86	GEARFIELD	CHICKEN	1.20
08	84	SHARPLEY	POWELL	90
09	225	DAVIS	ATLANTA	79
0b	92	STEFAN	STEFAN	160
07	180	LEAF	LEAF	245
08	147	PAUL	PAUL	180
09	236	SMITH	SMITH	274
0b	181	SMITH	SMITH	180
07	275	SMITH	SMITH	327
08	232	SMITH	SMITH	286
09	328	SMITH	SMITH	328
0b	287	SMITH	SMITH	407

RESULT	7
AB55988	
ID	AB55988 Standard, Protein: 595 AA.
XX	
XX	AB55988:
AC	
XX	
DT	26 MAR 2002 (11:57:00.77)
XX	
DE	Protophila melanogaster polyprophido 33921120.6756.

TS	CC	FEE	-GGG	GCGAGC	TSGA414
TR	24-FEB-2000:	260080-US04914.			
TR	24-FEB-2000:	260080-US05004.			
TR	01-MAR-2000:	260080-US05601.			
TR	20-MAR-2000:	260080-US07377.			
PR	21-MAR-2000:	260080-US07532.			
PE	16-MAY-2000:	260080-US08479.			
PR	17-MAY-2000:	260080-US13705.			
PT	22-MAY-2000:	260080-US14312.			
PR	30-MAY-2000:	260080-US14841.			
PR	02-JUN-2000:	260080-US15264.			
PR	10-JUN-2000:	260080-US16873.			
XX	(GENE) GENENTECH INC.				
XX	Baker KP, Berensini M, DeGeorge L, Desnoyers L, Filvaroff E, Gao W,				
EI	Gerritsen ME, Goldard A, Goldowski FJ, Kearney AL, Sherwood S;				
XX	Smith V, Stewart VA, Tumas D, Watanabe CK, Wood WJ, Zhang Z;				
XX	WPI: 2001-408281/43.				
CR	N USDB, AACSI304.				
FT	Isolated, secretory and transmembrane pro polypeptide used to detect				
PT	other pro polypeptides, link bioactive molecules to cells expressing				
EI	pro polypeptides, and detect the presence of mammalian lung, colon,				
PT	lung, breast, prostate, cervical				
XX	Claim 12; Fig 122; 81kbp; English.				
XX	AAM1272 AAC1246 represent covei human secretory and transmembrane				
CC	PRO polypeptides. The PRO polypeptides are useful to detect other				
CC	PRO polypeptides, via the bioactive molecules to cells expressing				
CC	PRO polypeptides, to modulate biological activities of cells expressing				
CC	PRO polypeptides, and to detect the presence of mammalian lung, colon,				
CC	breast, prostate, rectal, cervical or liver tumors by comparing pro				
CC	polypeptide expression in a cell sample to that in a control sample.				
CC	Some of the 275 sequences are also useful to stimulate the release of				
CC	tumor necrosis factor-alpha (TNF-alpha) from human blood, the				
CC	proliferation or differentiation of chondrocytes, the proliferation or				
CC	gene expression in pericyte cells, the release of proteolytic enzymes from				
CC	catalase, the proliferation of inner ear utricular support cells or				
CC	of T lymphocytes, the release of a cytokine from peripheral blood				
CC	monocytes (PBMCs), or the proliferation of endothelial cells. Some of				
CC	the PRO polypeptides may mediate glucose or free fatty acid uptake by				
CC	skeletal muscle cells or by adipocytes; or inhibit binding of A-protein				
CC	to factor VIIA. The PRO polypeptides can be used in assays to identify				
CC	molecules involved in binding interactions. The polynucleotides encoding				
CC	PRO polypeptides can be used to generate probes, antisense RNA/DNA,				
CC	transgenic or knock out animals and can be used in gene therapy.				
XX	Sequence 511 AA:				
SQ					
Query Match	4.8% Score 93.5 DB 22 Length 511				
Best Local Similarity	18.5% Prod No. 0.77				
Matches	77 Conservative 51 Mismatches 125 Indels 163 Gaps 20				
Y	23 IPEDSHLVKKGIDHLLPMWRRLSLPHLNSLYVGLASSLSOASCTKKENYLHELTGYQG R2				
LQ				
Y	83 GLIASASESDMDVR--GEERPSMGAAL-----YLIALKANFEVNRIRRG..... 124				
ZB	182 adstfntl eesqgslsmldvrlapgytvgllfltlsskyglftltmmmdlpz 248				
OY	125 -----DRIVSQIKWFLIEDKRRAIDTANAAGLAF---TCVKRSNPNNGRGRITMALRYV 175				
ZB	219 drvefrkrklhlwksatkaalpfoellcsevecdahdt fahetrtvrrttit 242				
OY	175 PEFILAKQTDFEGHGWNVSPLAIQELMSDPWFAPFMATKAQAVNLISLGCAFVNVA 245				
ZB	293 re-----hhdltpilpaee-kaaakirkqlqt..... 419				

[illegible]

```

15  AA07372 standard, Protein, 618 AA,
XX
AC  AA573921:
XX
16  29 MAY 2001 (first entry)
XX
DE  Rat NIS.
XX
FM  Rat; sodium iodide symporter; NIS; transmembrane expression;
KM  transgene localisation.
XX
OS  Rattus sp.
XX
PN  M0200113106-A1.
XX
FD  22 FEB 2001.
XX
XX  17-AUG-2000; 2000W0-US22566.
XX
XX  17-AUG-1999; 990S-0149168.
XX  16-AUG-2000; 2000US-0639667.
XX  16 AUG 2000; 2000US 0649198.
XX
PA  (MAYO-) MAYO FOUND MEDICAL EDUCATION 6 RES.
XX
P1  Russell SJ, Morris J, Peng K:
XX
XX  WU: 2001.257542/26.
XX  N-PSDB; AAF79537.
XX
PT  Monitoring transgene expression and therapeutic peptide production in
T7  mammals by detecting marker polypeptides linked to transgenes or
XX  therapeutic genes released from cells into extracellular body fluid -
XX  Disclosure: Fig 11; 79pp; English.
XX
XX  The present sequence is a self-cleaving linker. It may be used in a
XX  method for monitoring expression and/or localisation of a transgene, and
XX  production of therapeutic peptide in a mammal. The method involves
XX  quantifying or detecting the amount of marker polypeptide and/or sodium
XX  iodide symporter (NIS) linked to the product of the transgene or
XX  therapeutic gene released from cells into extracellular body fluid, or
XX  determining the location of labelled molecules which are transported into
XX  the cells bearing the marker peptide. The method provides a quick and
XX  effective monitoring of the level and kinetics of expression of
XX  transgenes and the tissue-specific distribution of expressed transgenes
XX  in cells, tissues, animals or humans without the need for disruptive and
XX  expensive sampling methods including surgery. The transgene location can
XX  be monitored without adversely affecting the mammal or the cell. The NIS
XX  is a self protein and as such does not stimulate a host immune
XX  reaction. Furthermore, the NIS functions solely to sequester iodine
XX  into a cell, which does not adversely affect normal cellular function
XX  or overall cell physiology.
XX
XX  Sequence 618 AA:
XX
XX  Query Match 4.88; Score 93.5; DB 22; Length 618;
XX  Host Local Similarity 25.28; Pred No. 1;
XX  Matches 609 Conservative 450 Mismatches 94; Indels 49; Gaps 14
XX
XX  97 GGGKPSM---GQALATVIALRANC---PEVGHKGRDVSQMKWLEDERRAI----- 142
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  Db 275 GGGKALIVGGIGLILVSAACQIVDYKKDCDPLIGVLSADYGYMPLIVIDL 333
XX
XX  143 -GTAAAGLAFCTGKRSNRRGSPLEETMA:ETVEEELFPAITGDIHGNVSTFATG 201
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  Db 334 EDIPVEVILIAVYGLSLASLSIMAAVAY EDIIPV 345
XX
XX  202 TMSMGCAET GTGCTAFATVLAIRGDMVGAALMSGL--TVLIRVYTLIRLFG 257
XX  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX  Db 386 I SKGLSTLYSACIEI AALSLTGGQVLTGGSTTGGVTSIPPLIGAIT-IGMT 439

```


0Y 258 CLAPRYMLEPAAETIPQTOEISVTIOV---LSLLPYROOSISVL---AGSIVEDVL 308
| | | | | : | | | | | : | | | | | : | | | | |
Db 440 cmppri:sgiaag--- larslwravagatlppqeqcmaglpctsaagctlaasrl 489

Search completed: September 4, 2002, 09:42:41
Job time: 6028 sec

•

•

•


```

Db 61 SSIAGGKEDLYLAKIAGYQGLISAFSEDEPWVQKPSWQAVILLALPANEVPR 120
QY 121 GHRRLAVSEI FRETFFPPAT----- 142
Db 121 GHRRLAVSEI FRETFFPPAT----- 142
QY 143 -----DTAMAGLAFGLKRSNENPQRIPTMALPVPPEELKAGTECHP 189
Db 181 YAVRPFHQHSVDTAAMAGLAFGLKRSNENPQRIPTMALPVPPEELKAGTECHP 240
QY 190 GNVSTPLALQELMTSPMPAPITGTAI KAPVALLASLQAPQNAI MTSQII PVI NEKT 249
Db 241 GNVSTPLALQELMTSPMPAPITGTAI KAPVALLASLQAPQNAI MTSQII PVI NEKT 249
QY 250 YTDILFPGKLABPMLEPAETTPQGPISVTLVSTLPPPPGISTVIANSTVAVIK 309
Db 301 YTDILFPGKLABPMLEPAETTPQGPISVTLVSTLPPPPGISTVIANSTVAVIK 360
QY 410 KAHLEAGPEYEVQASISAPVITSYMKAGREPEWALPPNPPILOGIAVPPKIDETI 369
Db 461 KAHLEAGPEYEVQASISAPVITSYMKAGREPEWALPPNPPILOGIAVPPKIDETI 420
QY 470 ELKLVSW 476
Db 421 ELKLVSW 427

```

RESULT 2

Transcobalamin 1 precursor [validated] - human

N:Alternative names: Haptocorrin; vitamin B12 binding protein, R binder family

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 08-Dec-2000

C:Accession: A34227; A42174; A60553

R:Johnson, J.; Hollkens, J.; Allen, R.H.; Berliner, N.

J. Biol. Chem. 264, 15754-15757, 1989

A:Title: Structure of the cDNA encoding transcobalamin 1, a neutrophil granule protein.

A:Reference number: A34227; M01D:89480156

A:Accession: A34227

A:Molecule type: mRNA

A:Residues: 1-433 <OH>

A:Cross references: DB:105068; M01D:907478; P01N:AAAG1058.1; P01D:907479

R:Johnson, J.; Yang-Peng, T.; Berliner, N.

Genomics 12, 459-464, 1992

A:Title: Genome structure and mapping of the chromosomal gene for transcobalamin 1 (TCN)

A:Reference number: A42174; M01D:92217066

A:Accession: A42174

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-26 <OH>

A:Note: Sequence extracted from NCR1 backbone (NCBIN:95735; NCBIp:95736)

R:Johnson, P.; Thim, L.; Noxo, E.

Stand. J. Clin. Lab. Invest. 50, 187-194, 1990

A:Title: Epidermal growth factor and haptocorrin in nasal secretion.

A:Reference number: A60553; M01D:90251988

A:Accession: A60553

A:Molecule type: protein

A:Residues: 24-25, 'X', 27-32, 'X', 34 38, 'X', 40 <OH>

A:Gene: GDB:TCN1

A:Cross references: GDB:118882; OMIM:189905

A:Mp position: 11911-11912.3

C:Superfamily: gastrin intrinsc factor

E:1-2/3/4/5/6/7/8/9/10/11/12/13/14/15/16/17/18/19/20/21/22/23/24/25/26/27/28/29/30/31/32/33/34/35/36/37/38/39/40/41/42/43/44/45/46/47/48/49/50/51/52/53/54/55/56/57/58/59/60/61/62/63/64/65/66/67/68/69/69/70/71/72/73/74/75/76/77/78/79/80/81/82/83/84/85/86/87/88/89/90/91/92/93/94/95/96/97/98/99/100/101/102/103/104/105/106/107/108/109/110/111/112/113/114/115/116/117/118/119/120/121/122/123/124/125/126/127/128/129/130/131/132/133/134/135/136/137/138/139/140/141/142/143/144/145/146/147/148/149/150/151/152/153/154/155/156/157/158/159/160/161/162/163/164/165/166/167/168/169/170/171/172/173/174/175/176/177/178/179/180/181/182/183/184/185/186/187/188/189/190/191/192/193/194/195/196/197/198/199/200/201/202/203/204/205/206/207/208/209/210/211/212/213/214/215/216/217/218/219/220/221/222/223/224/225/226/227/228/229/230/231/232/233/234/235/236/237/238/239/240/241/242/243/244/245/246/247/248/249/250/251/252/253/254/255/256/257/258/259/260/261/262/263/264/265/266/267/268/269/270/271/272/273/274/275/276/277/278/279/280/281/282/283/284/285/286/287/288/289/290/291/292/293/294/295/296/297/298/299/300/301/302/303/304/305/306/307/308/309/310/311/312/313/314/315/316/317/318/319/320/321/322/323/324/325/326/327/328/329/330/331/332/333/334/335/336/337/338/339/340/341/342/343/344/345/346/347/348/349/350/351/352/353/354/355/356/357/358/359/360/361/362/363/364/365/366/367/368/369/370/371/372/373/374/375/376/377/378/379/380/381/382/383/384/385/386/387/388/389/390/391/392/393/394/395/396/397/398/399/400/401/402/403/404/405/406/407/408/409/410/411/412/413/414/415/416/417/418/419/420/421/422/423/424/425/426/427/428/429/430/431/432/433/434/435/436/437/438/439/440/441/442/443/444/445/446/447/448/449/450/451/452/453/454/455/456/457/458/459/460/461/462/463/464/465/466/467/468/469/470/471/472/473/474/475/476/477/478/479/480/481/482/483/484/485/486/487/488/489/490/491/492/493/494/495/496/497/498/499/500/501/502/503/504/505/506/507/508/509/510/511/512/513/514/515/516/517/518/519/520/521/522/523/524/525/526/527/528/529/530/531/532/533/534/535/536/537/538/539/540/541/542/543/544/545/546/547/548/549/550/551/552/553/554/555/556/557/558/559/560/561/562/563/564/565/566/567/568/569/570/571/572/573/574/575/576/577/578/579/580/581/582/583/584/585/586/587/588/589/590/591/592/593/594/595/596/597/598/599/600/601/602/603/604/605/606/607/608/609/610/611/612/613/614/615/616/617/618/619/620/621/622/623/624/625/626/627/628/629/630/631/632/633/634/635/636/637/638/639/640/641/642/643/644/645/646/647/648/649/650/651/652/653/654/655/656/657/658/659/660/661/662/663/664/665/666/667/668/669/670/671/672/673/674/675/676/677/678/679/680/681/682/683/684/685/686/687/688/689/690/691/692/693/694/695/696/697/698/699/700/701/702/703/704/705/706/707/708/709/710/711/712/713/714/715/716/717/718/719/720/721/722/723/724/725/726/727/728/729/730/731/732/733/734/735/736/737/738/739/740/741/742/743/744/745/746/747/748/749/750/751/752/753/754/755/756/757/758/759/760/761/762/763/764/765/766/767/768/769/770/771/772/773/774/775/776/777/778/779/780/781/782/783/784/785/786/787/788/789/790/791/792/793/794/795/796/797/798/799/800/801/802/803/804/805/806/807/808/809/810/811/812/813/814/815/816/817/818/819/820/821/822/823/824/825/826/827/828/829/830/831/832/833/834/835/836/837/838/839/840/841/842/843/844/845/846/847/848/849/850/851/852/853/854/855/856/857/858/859/860/861/862/863/864/865/866/867/868/869/870/871/872/873/874/875/876/877/878/879/880/881/882/883/884/885/886/887/888/889/890/891/892/893/894/895/896/897/898/899/900/901/902/903/904/905/906/907/908/909/910/911/912/913/914/915/916/917/918/919/920/921/922/923/924/925/926/927/928/929/930/931/932/933/934/935/936/937/938/939/940/941/942/943/944/945/946/947/948/949/950/951/952/953/954/955/956/957/958/959/960/961/962/963/964/965/966/967/968/969/970/971/972/973/974/975/976/977/978/979/980/981/982/983/984/985/986/987/988/989/990/991/992/993/994/995/996/997/998/999/1000/1001/1002/1003/1004/1005/1006/1007/1008/1009/1010/1011/1012/1013/1014/1015/1016/1017/1018/1019/1020/1021/1022/1023/1024/1025/1026/1027/1028/1029/1030/1031/1032/1033/1034/1035/1036/1037/1038/1039/1040/1041/1042/1043/1044/1045/1046/1047/1048/1049/1050/1051/1052/1053/1054/1055/1056/1057/1058/1059/1060/1061/1062/1063/1064/1065/1066/1067/1068/1069/1070/1071/1072/1073/1074/1075/1076/1077/1078/1079/1080/1081/1082/1083/1084/1085/1086/1087/1088/1089/1090/1091/1092/1093/1094/1095/1096/1097/1098/1099/1100/1101/1102/1103/1104/1105/1106/1107/1108/1109/1110/1111/1112/1113/1114/1115/1116/1117/1118/1119/1120/1121/1122/1123/1124/1125/1126/1127/1128/1129/1130/1131/1132/1133/1134/1135/1136/1137/1138/1139/1140/1141/1142/1143/1144/1145/1146/1147/1148/1149/1150/1151/1152/1153/1154/1155/1156/1157/1158/1159/1160/1161/1162/1163/1164/1165/1166/1167/1168/1169/1170/1171/1172/1173/1174/1175/1176/1177/1178/1179/1180/1181/1182/1183/1184/1185/1186/1187/1188/1189/1190/1191/1192/1193/1194/1195/1196/1197/1198/1199/1200/1201/1202/1203/1204/1205/1206/1207/1208/1209/1210/1211/1212/1213/1214/1215/1216/1217/1218/1219/1220/1221/1222/1223/1224/1225/1226/1227/1228/1229/1230/1231/1232/1233/1234/1235/1236/1237/1238/1239/1240/1241/1242/1243/1244/1245/1246/1247/1248/1249/1250/1251/1252/1253/1254/1255/1256/1257/1258/1259/1260/1261/1262/1263/1264/1265/1266/1267/1268/1269/1270/1271/1272/1273/1274/1275/1276/1277/1278/1279/1280/1281/1282/1283/1284/1285/1286/1287/1288/1289/1290/1291/1292/1293/1294/1295/1296/1297/1298/1299/1300/1301/1302/1303/1304/1305/1306/1307/1308/1309/1310/1311/1312/1313/1314/1315/1316/1317/1318/1319/1320/1321/1322/1323/1324/1325/1326/1327/1328/1329/1330/1331/1332/1333/1334/1335/1336/1337/1338/1339/1340/1341/1342/1343/1344/1345/1346/1347/1348/1349/1350/1351/1352/1353/1354/1355/1356/1357/1358/1359/1360/1361/1362/1363/1364/1365/1366/1367/1368/1369/1370/1371/1372/1373/1374/1375/1376/1377/1378/1379/1380/1381/1382/1383/1384/1385/1386/1387/1388/1389/1390/1391/1392/1393/1394/1395/1396/1397/1398/1399/1400/1401/1402/1403/1404/1405/1406/1407/1408/1409/1410/1411/1412/1413/1414/1415/1416/1417/1418/1419/1420/1421/1422/1423/1424/1425/1426/1427/1428/1429/1430/1431/1432/1433/1434/1435/1436/1437/1438/1439/1440/1441/1442/1443/1444/1445/1446/1447/1448/1449/1450/1451/1452/1453/1454/1455/1456/1457/1458/1459/1460/1461/1462/1463/1464/1465/1466/1467/1468/1469/1470/1471/1472/1473/1474/1475/1476/1477/1478/1479/1480/1481/1482/1483/1484/1485/1486/1487/1488/1489/1490/1491/1492/1493/1494/1495/1496/1497/1498/1499/1500/1501/1502/1503/1504/1505/1506/1507/1508/1509/1510/1511/1512/1513/1514/1515/1516/1517/1518/1519/1520/1521/1522/1523/1524/1525/1526/1527/1528/1529/1530/1531/1532/1533/1534/1535/1536/1537/1538/1539/1540/1541/1542/1543/1544/1545/1546/1547/1548/1549/1550/1551/1552/1553/1554/1555/1556/1557/1558/1559/1560/1561/1562/1563/1564/1565/1566/1567/1568/1569/1570/1571/1572/1573/1574/1575/1576/1577/1578/1579/1580/1581/1582/1583/1584/1585/1586/1587/1588/1589/1590/1591/1592/1593/1594/1595/1596/1597/1598/1599/1600/1601/1602/1603/1604/1605/1606/1607/1608/1609/1610/1611/1612/1613/1614/1615/1616/1617/1618/1619/1620/1621/1622/1623/1624/1625/1626/1627/1628/1629/1630/1631/1632/1633/1634/1635/1636/1637/1638/1639/1640/1641/1642/1643/1644/1645/1646/1647/1648/1649/1650/1651/1652/1653/1654/1655/1656/1657/1658/1659/1660/1661/1662/1663/1664/1665/1666/1667/1668/1669/1670/1671/1672/1673/1674/1675/1676/1677/1678/1679/1680/1681/1682/1683/1684/1685/1686/1687/1688/1689/1690/1691/1692/1693/1694/1695/1696/1697/1698/1699/1700/1701/1702/1703/1704/1705/1706/1707/1708/1709/1710/1711/1712/1713/1714/1715/1716/1717/1718/1719/1720/1721/1722/1723/1724/1725/1726/1727/1728/1729/1730/1731/1732/1733/1734/1735/1736/1737/1738/1739/1740/1741/1742/1743/1744/1745/1746/1747/1748/1749/1750/1751/1752/1753/1754/1755/1756/1757/1758/1759/1760/1761/1762/1763/1764/1765/1766/1767/1768/1769/1770/1771/1772/1773/1774/1775/1776/1777/1778/1779/1780/1781/1782/1783/1784/1785/1786/1787/1788/1789/1790/1791/1792/1793/1794/1795/1796/1797/1798/1799/1800/1801/1802/1803/1804/1805/1806/1807/1808/1809/1810/1811/1812/1813/1814/1815/1816/1817/1818/1819/1820/1821/1822/1823/1824/1825/1826/1827/1828/1829/1830/1831/1832/1833/1834/1835/1836/1837/1838/1839/1840/1841/1842/1843/1844/1845/1846/1847/1848/1849/1850/1851/1852/1853/1854/1855/1856/1857/1858/1859/1860/1861/1862/1863/1864/1865/1866/1867/1868/1869/1870/1871/1872/1873/1874/1875/1876/1877/1878/1879/1880/1881/1882/1883/1884/1885/1886/1887/1888/1889/1890/1891/1892/1893/1894/1895/1896/1897/1898/1899/1900/1901/1902/1903/1904/1905/1906/1907/1908/1909/1910/1911/1912/1913/1914/1915/1916/1917/1918/1919/1920/1921/1922/1923/1924/1925/1926/1927/1928/1929/1930/1931/1932/1933/1934/1935/1936/1937/1938/1939/1940/1941/1942/1943/1944/1945/1946/1947/1948/1949/1950/1951/1952/1953/1954/1955/1956/1957/1958/1959/1960/1961/1962/1963/1964/1965/1966/1967/1968/1969/1970/1971/1972/1973/1974/1975/1976/1977/1978/1979/1980/1981/1982/1983/1984/1985/1986/1987/1988/1989/1990/1991/1992/1993/1994/1995/1996/1997/1998/1999/2000/2001/2002/2003/2004/2005/2006/2007/2008/2009/2010/2011/2012/2013/2014/2015/2016/2017/2018/2019/2020/2021/2022/2023/2024/2025/2026/2027/2028/2029/2030/2031/2032/2033/2034/2035/2036/2037/2038/2039/2040/2041/2042/2043/2044/2045/2046/2047/2048/2049/2050/2051/2052/2053/2054/2055/2056/2057/2058/2059/2060/2061/2062/2063/2064/2065/2066/2067/2068/2069/2070/2071/2072/2073/2074/2075/2076/2077/2078/2079/2080/2081/2082/2083/2084/2085/2086/2087/2088/2089/2090/2091/2092/2093/2094/2095/2096/2097/2098/2099/2100/2101/2102/2103/2104/2105/2106/2107/2108/2109/2110/2111/2112/2113/2114/2115/2116/2117/2118/2119/2120/2121/2122/2123/2124/2125/2126/2127/2128/2129/2130/2131/2132/2133/2134/2135/2136/2137/2138/2139/2140/2141/2142/2143/2144/2145/2146/2147/2148/2149/2150/2151/2152/2153/2154/2155/2156/2157/2158/2159/2160/2161/2162/2163/2164/2165/2166/2167/2168/2169/2170/2171/2172/2173/2174/2175/2176/2177/2178/2179/2180/2181/2182/2183/2184/2185/2186/2187/2188/2189/2190/2191/2192/2193/2194/2195/2196/2197/2198/2199/2200/2201/2202/2203/2204

Matches 72: Conservative 47: Mismatches 87: Indels 115: Gaps 18:

QY GSASHPN-----GPGGKNSKQAVYLAIPANCPVP-GRPG 124
 DB GSASHPN-----GPGGKNSKQAVYLAIPANCPVP-GRPG 124

QY 125 PRVVSATKMTF--PPVATITAMACT--ATGCTKPSNTHGPPGATTAATVEPEI 179
 DB 125 PRVVSATKMTF--PPVATITAMACT--ATGCTKPSNTHGPPGATTAATVEPEI 179

QY 146 PQAITE-----VLASDFG-----NSAHLHLSGHTIYNNRV 179
 DB 146 PQAITE-----VLASDFG-----NSAHLHLSGHTIYNNRV 179

QY 180 LKFAIPGHPVNYSTFAIPELMSIPGAEELACLAFAVALASVGCAT--GNA 245
 DB 180 LKFAIPGHPVNYSTFAIPELMSIPGAEELACLAFAVALASVGCAT--GNA 245

QY 246 LMSQILPVNPK ITIDLPGLARVWLEPAETIP-----Q 274
 DB 246 LMSQILPVNPK ITIDLPGLARVWLEPAETIP-----Q 274

QY 275 T-QGLTIVTIVGLST--PPVPSISVLAISIVE--GVKFAHE-LGGPEYELASISG 327
 DB 275 T-QGLTIVTIVGLST--PPVPSISVLAISIVE--GVKFAHE-LGGPEYELASISG 327

QY 277 TIGTASVGYPTVIGYELSPKRRGVPSVYITFTFYVAVGAMALVY--AG 285
 DB 277 TIGTASVGYPTVIGYELSPKRRGVPSVYITFTFYVAVGAMALVY--AG 285

QY 428 PY-----LISYWGKAA 338
 DB 428 PY-----LISYWGKAA 338

QY 286 PFKSSFNADMLLASVGRKAS 306
 DB 286 PFKSSFNADMLLASVGRKAS 306

RESULT 9

H72031
 I:Protein acid syn:betase (imported) Chlamydia pneumoniae (serotype TW029 and
 C:Species: Chlamydia pneumoniae Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23 Apr 1999 #text_change 11 May 2000
 C:Accession: H72031: C81509
 K:Kalam, S.; Mitchell, W.; Marathe, K.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21: 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000: MIMD:9920606
 A:Accession: H72031
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1407 <NN>
 A:Cross-references: GB:AE001663; GB:AE001343; NID:94177134; PID:AA018969.1; PID:9437714
 A:Experimental source: strain TW029
 A:Head: T.D.; Hincham, R.C.; Shew, G.; Gill, S.R.; Boddle-Perera, J.P.; Miller, G.; Hickey,
 C.; Dodson, E.; Swinn, M.; Nelson, W.; Levey, R.; Kolonay, J.; McQuarrie, G.; Saliberry,
 Nucleic Acids Res. 28: 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis and Chlamydia pneumoniae
 A:Reference number: A81500: MIMD:20150255
 A:Accession: C81509
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1407 <NN>
 A:Cross-references: GB:AF009913; GB:AF009913; NID:97189950; PID:AA039813.1; PID:9718995
 A:Experimental source: strain AR39, HL cells
 A:Genetics:
 A:Gene: lmpA; CP1038
 A:Superfamily: lmpA and syn:bet

Query Match

Best Local Similarity 5.48: Score 104.5; DB 2: Length 307;
 Matches 72: Conservative 47: Mismatches 87: Indels 115: Gaps 18;

QY GSASHPN-----GPGGKNSKQAVYLAIPANCPVP-GRPG 124
 DB GSASHPN-----GPGGKNSKQAVYLAIPANCPVP-GRPG 124

QY 125 PRVVSATKMTF--PPVATITAMACT--ATGCTKPSNTHGPPGATTAATVEPEI 179
 DB 125 PRVVSATKMTF--PPVATITAMACT--ATGCTKPSNTHGPPGATTAATVEPEI 179

QY 146 PQAITE-----VLASDFG-----NSAHLHLSGHTIYNNRV 179
 DB 146 PQAITE-----VLASDFG-----NSAHLHLSGHTIYNNRV 179

QY 180 LKFAIPGHPVNYSTFAIPELMSIPGAEELACLAFAVALASVGCAT--GNA 245
 DB 180 LKFAIPGHPVNYSTFAIPELMSIPGAEELACLAFAVALASVGCAT--GNA 245

DB 146 PQAITE-----VLASDFG-----NSAHLHLSGHTIYNNRV 179
 QY 246 LMSQILPVNPK ITIDLPGLARVWLEPAETIP-----Q 274
 DB 246 LMSQILPVNPK ITIDLPGLARVWLEPAETIP-----Q 274

QY 275 T-QGLTIVTIVGLST--PPVPSISVLAISIVE--GVKFAHE-LGGPEYELASISG 327
 DB 275 T-QGLTIVTIVGLST--PPVPSISVLAISIVE--GVKFAHE-LGGPEYELASISG 327

QY 277 TIGTASVGYPTVIGYELSPKRRGVPSVYITFTFYVAVGAMALVY--AG 285
 DB 277 TIGTASVGYPTVIGYELSPKRRGVPSVYITFTFYVAVGAMALVY--AG 285

QY 428 PY-----LISYWGKAA 338
 DB 428 PY-----LISYWGKAA 338

QY 286 PFKSSFNADMLLASVGRKAS 306
 DB 286 PFKSSFNADMLLASVGRKAS 306

RESULT 10

G96749
 I:Protein acid syn:betase (imported) Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02 Mar 2001 #text_change 31 Mar 2001
 C:Accession: G96749
 K:Chen, G.W.; Chang, M.K.; Gann, L.; Conway, A.B.; Conway, A.B.; Crossley, T.H.; Dewar,
 Nature 408: 816-820, 2000
 A:Author: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 G.; Liu, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lopez, J.S.; Maitl, R.; Marzita
 F.; May, M.; Kennedy, T.; Kowalev, D.; Sakano, H.
 A:Author: Sakurai, S.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talbot,
 Lee, M.; Wu, D.; Yu, D.; Fraser, G.M.; Walker, J.D.; Davies, B.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A8141: MIMD:2016719
 A:Accession: G96749
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1450 <STOS>
 A:Cross-references: GB:AE005173; NID:96648174; PID:AA021174.1; GSPDB:GN00141
 A:Genetics:
 A:Gene: P28P22.27
 A:Map position: 1

Query Match

Best Local Similarity 5.28: Score 109; DB 2: Length 450;
 Matches 56: Conservative 27: Mismatches 71: Indels 90: Gaps 11;

QY 125 PRVVSATKMTF--PPVATITAMACT--ATGCTKPSNTHGPPGATTAATVEPEI 179
 DB 125 PRVVSATKMTF--PPVATITAMACT--ATGCTKPSNTHGPPGATTAATVEPEI 179

QY 146 PQAITE-----VLASDFG-----NSAHLHLSGHTIYNNRV 179
 DB 146 PQAITE-----VLASDFG-----NSAHLHLSGHTIYNNRV 179

QY 180 LKFAIPGHPVNYSTFAIPELMSIPGAEELACLAFAVALASVGCAT--GNA 245
 DB 180 LKFAIPGHPVNYSTFAIPELMSIPGAEELACLAFAVALASVGCAT--GNA 245

QY 286 PFKSSFNADMLLASVGRKAS 306
 DB 286 PFKSSFNADMLLASVGRKAS 306

QY 428 PY-----LISYWGKAA 338
 DB 428 PY-----LISYWGKAA 338

QY 246 LMSQILPVNPK ITIDLPGLARVWLEPAETIP-----Q 274
 DB 246 LMSQILPVNPK ITIDLPGLARVWLEPAETIP-----Q 274

QY 275 T-QGLTIVTIVGLST--PPVPSISVLAISIVE--GVKFAHE-LGGPEYELASISG 327
 DB 275 T-QGLTIVTIVGLST--PPVPSISVLAISIVE--GVKFAHE-LGGPEYELASISG 327

QY 277 TIGTASVGYPTVIGYELSPKRRGVPSVYITFTFYVAVGAMALVY--AG 285
 DB 277 TIGTASVGYPTVIGYELSPKRRGVPSVYITFTFYVAVGAMALVY--AG 285

QY 428 PY-----LISYWGKAA 338
 DB 428 PY-----LISYWGKAA 338

QY 286 PFKSSFNADMLLASVGRKAS 306
 DB 286 PFKSSFNADMLLASVGRKAS 306

RESULT 11

A48126
 I:Transcription activator GSN1 - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 21-Jan-1991 #sequence_revision 18 Nov 1991 #text_change 21 Jan 2000

DB 541 HVATKISQHEVEVLDQJFO... NIEKPF TSPECKNLSTKVYDWLIRPIETO 589

OY 454 LLO 356

DB 590 LSO 592

Search completed: September 4, 2002, 09:41:19
Job time: 6851 sec

[illegible][illegible]

[illegible]

AC005006	1998	101720 bp	DNA	linear	Feb 03-Jul-2000
DEFINITION	Homo sapiens clone RP1-56J10, complete sequence.				
ACCESSION	AC005006				
VERSION	AC005006.2 GI:5406405				
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eskayota, Melissa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eulalia; Primates; Catarrhini; Homidae; Homo.				
AUTHORS	1 (bases 1 to 101720)				
TITLE	The sequence of Homo sapiens clone				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 101720)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (12 Jun 1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	3 (bases 1 to 101720)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (14 Jan 1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	4 (bases 1 to 101720)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (01 Jun 1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	5 (bases 1 to 101720)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (03 Jun 1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	6 (bases 1 to 101720)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (19 Jun 1999) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
REFERENCE	7 (bases 1 to 101720)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (21 Dec 1999) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
REFERENCE	8 (bases 1 to 101720)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (03 Jun 2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA				
COMMENT	On Jul 1, 1999 this sequence version replaced GI:4156152. Center project name: H.DJ56J10.				
FEATURES	Location/Qualifiers				
source	1..101720				
ORGANISM	"Homo sapiens"				
db_xref	"taxon:9606"				
clone	"RP1-56J10"				
BASE COUNT	24693 A 26400 C 25564 G 24063 T				
ORIGIN					
Query Match	34.6% Score 636.4; E=0; Length 101720;				
Best local Similarity	99.8%; Pred. No. 2,76-143;				
Matches 737, Conserved 170	0.8% Similes 1; Indels 0; Gaps 0;				
Db 75740	AGGATTATGATCTTACATCTAATCAATGAAATCAATATATATCTGATCTTATCTG 75740				

JOURNAL Submitted (29 JAN 1999) Protein Chemistry Laboratory, Dept

Molecular and Structural Biology, University of Aarhus, 8000

Weds 10 C, Aarhus 8000, Denmark

FEATURES

Source

Local job/Qualifiers

1..1299

/organism "Hos. taurus"

/db_xref "taxon:9913"

1..1299

/cds_start=1

/product "transcobalamin"

/protein_id "AAU23829.1"

/db_xref "GI:4572454"

/translation "MGIHGLIFLLIGLALNCEITFVDSILVERIGORLPMR

LSOHLNPIYVERISSEZAAFTAMTHSLFVSSZLFAASKEELIABEEM

60

60ATVITLAKANDFFISGRNDELVSLEKELERKALIDNKHRIISYVSLG

110

110ALCVHOFVHESVCKELVAVERHRLKLVSVLIMAMAKALISLILNATG

120

120PHEATIKRVAERIKKQTPFEGSNVSTELALILMSIPSEVSLGAKAKA

130

130LASTGKTPQNDIMSLTPEVNGKSVYDLSRQDVAITPEALFTPEAVVEE

140

140IIVLKVSTISSESTHSSVHAGSSLEHLEKNAGHCHREFRRCASLSGPIITSVLR

150

150KAPPEPWVTPPPPPIGGTAAYPPPEDETLEGLVGM"

160

160

170

595

600

605

610

615

620

625

630

635

640

645

650

655

660

665

670

675

680

685

690

695

700

705

710

715

720

725

730

735

740

745

750

755

760

765

770

775

780

785

790

795

800

805

810

815

820

825

830

835

840

845

850

855

860

865

870

875

880

885

890

895

900

905

910

915

920

925

930

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

1..1299

Template: 10 ng
 Primer: each 5 pm
 dNTPs: each 4 mM
 Taq polymerase: 0.025 units/ul
 Total Vol: 20 ul

But let's

Mg:1.2: 1.5 mm

KCl: 50 mM

Tris-HCl: 10 mM

pH: 9.3

FEATURING

501170

Location/Qualifiers

1. 349

“Organism” = “Homo sapiens”

```

/db_xret = "taxon:9606"

```

$$/\text{map} = "768_(\text{B}, \text{C})_7"$$

514

for inner_bond

for linear_bind

CONCLUSION

ORIGIN

```

d      complement(300, .319)
68 a    110 c    101 q    70 t

```

Query Match

Best Local

99.78

Pred. No. 5.4e-79;

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Matches 34

tive

U; Mismatch

Indels

0; Gaps

1232 gataccccagcacaagccctcgagggccctatāccācālgccāccllygāyāgāyā 1351

DB 1 GATCACCCTACCTACAAATCTTTGAAATCTTATATCATATGCTTGAATCTTGGAGCAAGAGA GCG

QY 1352 qccaagcatcttcctggaagtcttctgccaagtcgtgcctgcagt 1411

Idb 61 GCCAAGCATCTCCCTGGGAAGTCTTTCTGGCAACTTGTACCTATGTCCTCAGGCT 12

1412 CTTCCATGAAAGCCACCCTATGATCTGATGCGCATGAAACCATCTCAGGACTCCTTACCAAA 1471

BH-121 CTCCATGACCGCACTTAAAGCATTGGTAAGAAGGAATAAACAATTTGTGCCTTCAG

[illegible]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

Search completed: September 4, 2002, 06:35:49
Job time: 10675 sec

[illegible]

CC	influenza, heat or central nervous system diseases; detecting
CC	susceptibility to microbial infection; treating or preventing such
CC	diseases; forensic analysis; gene therapy; paternity testing; mapping
CC	genomic loci associated with phenotypic traits (and subsequent cloning
CC	of the genes responsible); and the production of transgenic organisms.
CC	Antibodies raised against (1) are useful as diagnostic and therapeutic
CC	tools and in drug screening; antibody AAH87644 represent the human
CC	DNA sequences containing biallelic polymorphic sites described in the
CC	invention.
XX	
XX	Sequence 334 BP: 65 A; 106 C; 95 G; 68 T; 0 other:
XX	
XX	Query Match 19.3% Score 332.4; MH 20; length 334;
XX	Best Local Similarity 29.7%; E-Val No. 1.9e-74;
XX	Matches 3337 Conservative 0; Mismatches 1; Indels 0; Gaps 0;
YY	1202 gattcagccgacggcgatgccttccttgaaggtgcatacagatggcgaccttgaggcagaga 1351
DB	1 gatcacccggacacaagactcttagaggtccctatacaatgajjaccan.tlgaagagaga 60
YY	1352 gccagacattcttcctgaggaaactttcttgcagaagctctgcacgactgagccctcagagt 1411
DB	61 gccacgacatctctctcttgcaggaactcttctctgccaatcttgcacacctctggccccagag 120
YY	1412 cctcccat taagtgatgcagat tttatgatgagtgatgaat ctctcagagatccttggaaa 1471
DB	121 cctcccataagagctccacccaatggtctcgal ggccttaaacatctcagactctcttgcaca 180
YY	1472 aaacagatctccgagagcagatgatatcttat gaadacatctctctctatgattaaatctccg 1531
DB	181 aaacagatctccgagagcagatgatatgtagagacacatctctctatgattaaatctccg 240
YY	1532 caaagaaatctctctcagacagagatataag actgaacacatctcacttctatt 1591
DB	241 caaagaaatctctctcagacagagatataag acacacacacacacacacacacacacacac 300
YY	1592 agaatgttagctctgaatctgtatctgtatctgtatctgtatctgtatctgtatctgtatct 1655
DB	301 agaatgttagctctgaatctgtatctgtatctgtatctgtatctgtatctgtatctgtatct 360
II	301 agaatgttagctctgaatctgtatctgtatctgtatctgtatctgtatctgtatctgtatct 364
RESULT	7
AAH87616	
ID	AAH87616 standard) DNA; 334 bp.
XX	
AC	AAH87616;
LI	27-FEB-2002 (first entry)
XX	
XX	Human single nucleotide polymorphism containing DNA sequence #2473.
XX	
KW	Biallelic marker; polymorphism; human disease diagnosis; treatment;
KW	phenotypic trait; gene therapy; forensic; paternity mapping; cancer;
KW	transgenic; single nucleotide polymorphism; SNP; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key Location/Qualifiers
FT	Variation replace(184..)
TI	/start ..end
FT	/standard_name="single nucleotide polymorphism"
DN	W0995.095.A2.
XX	
PD	21-OCT-1999.
XX	
PF	30-MAR-1999; 99M-US06893.
XX	
PR	09-APR-1998; 98US-0057871.
XX	
TA	(WHED) WHITEHEAD INST BIOMEDICAL RES.
XX	


```

XX
SQ Sequence 405 BP; 59 A; 44 C; 89 G; 68 T; 1 other:

Query Match: 11.1%; Score 141; Pos 21; Length 205;
Best Local Similarity: 99.5%; Prod. No. 1,28-382;
Matches 141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

UY 1 ttctactgctcaccacacacgtgctgctgagacacgttggaggtctctctctt 60
   |||
DB 114 ttgctacgtctaccacacacgtgctgctgagacacgttggaggtctctctctt 173
   |||

UY 61 ctggaggtctctctctctctctctctctctctctctctctctctctctctct 120
   |||
DB 174 ctggaggtctctctctctctctctctctctctctctctctctctctctctct 243
   |||

UY 121 gtaaaagat+ggggggagacacacacacacacacacacacacacacacacac 180
   |||
DB 234 gtaaaagatgagacacacacacacacacacacacacacacacacacacacacac 293
   |||

UY 181 aacccacacac 192
   |||
DB 234 aacccacacac 405
   |||

RESULT 14
ABR44337
ID ABR44337 standard; DNA; 586 BP.
AC ABR44337;
XX
XX
DI 01 FEB-2002 (first entry)
DE Human foetal liver single exon nucleic acid probe #12642.
XX
XX
FW Human foetal liver gene expression; single exon nucleic acid probe #12642.
XX
OS Homo sapiens.
XX
PN W0200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-0500669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633665.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000US-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX
WP1: 2001-AB3447/52
XX
XX
PI Human genome-derived single exon nucleic acid probes useful for
PI analyzing gene expression in human foetal liver
XX
XX
XX
Claim 1: SEQ ID NO 12642; 639pp - sequence listing; English.
XX
XX
The invention relates to a single exon nucleic acid probe for
measuring human gene expression in a sample derived from human foetal
liver. The single exon nucleic acid probes may be used for predicting,
measuring and displaying gene expression in samples derived from human
foetal liver. The present sequence is a single exon nucleic acid
probe of the invention.
XX
XX
Note: The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WHO at http://www.int.fdb/fdb/aligned\_pat\_sequences.

```

```

XX
SQ Sequence 586 BP; 128 A; 165 C; 160 G; 133 T; 0 other:

Query Match: 11.0%; Score 160; Pos 22; Length 586;
Best Local Similarity: 100.0%; Prod. No. 2,86-78;
Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

UY 628 caatctctctctctctctctctctctctctctctctctctctctctctct 587
   |||
DB 193 caatctctctctctctctctctctctctctctctctctctctctctctct 252
   |||

UY 688 aaggttctctctctctctctctctctctctctctctctctctctctctct 747
   |||
DB 253 aaggttctctctctctctctctctctctctctctctctctctctctctct 312
   |||

UY 748 caatctctctctctctctctctctctctctctctctctctctctctctct 807
   |||
DB 413 caatctctctctctctctctctctctctctctctctctctctctctctct 272
   |||

UY 808 gcaacacacac 817
   |||
DB 473 gcaacacacac 482
   |||

RESULT 14
ABR41474
ID ABR41474 standard; DNA; 586 BP.
AC ABR41474;
XX
XX
DI 23 JAN-2002 (first entry)
DE Probe #0940 for gene expression analysis in human heart cell sample.
XX
XX
FW Human gene expression; heart; microarray; vascular system; probe;
cardiovascular disease; hypertension; cardiac arrhythmia;
congenital heart disease; ss.
XX
XX
OS Homo sapiens.
XX
PN W0200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30 JAN 2001; 2001WO-0500669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 29-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633665.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000US-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX
WP1: 2001-AB3447/53.
XX
XX
PI Single exon nucleic acid probes for analyzing gene expression in human
PI hearts -
XX
XX
Claim 1: SEQ ID NO 940; 599pp; English.
XX
XX
The present invention relates to single exon nucleic acid probes for
measuring human gene expression in a sample derived from human heart. The
present sequence is one such probe. The probes may be used for
predicting, measuring and displaying gene expression in samples derived
from the human heart via microarrays. By measuring gene expression, the
probes are useful for predicting, diagnosis, grading, staging,
monitoring and prognostic diseases of the human heart and vascular system

```


CC cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/patlib/shed_pat_sequences.

XX
SQ Sequence 586 BP; 128 A; 165 C; 160 G; 133 T; 0 other;

Query Match 11.0%; Score 190; DB 22; Length 586;

Best Local Similarity 100.0%; Pred. No. 2,86-38;

Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 628 cagcttcacagacttcccccagccttgggagagatgggagagagagagagagagagag 687

DB 134 cagcttcacagacttcccccagccttgggagagatgggagagagagagagagagag 252

QY 668 aggttgccttgcctggccagcttgcagatggagagcttccagaatgctctcatgattcc 747

DB 273 aggttgccttgcctggccagcttgcagatggagagcttccagaatgctctcatgattcc 312

QY 748 cagcttgccttgcctggccagcttgcagatggagagcttccagaatgctctcatgattcc 807

DB 313 cagcttgccttgcctggccagcttgcagatggagagcttccagaatgctctcatgattcc 372

QY 808 gcaccacagag 817

DB 373 gcaccacagag 382

RESULT 15

AAK12801

AAK12801 standard; DNA; 586 BP.

AAK12801:

05-NOV-2001 (first entry)

Human brain expressed single exon probe SEQ ID NO: 12792.

Human; brain expressed exon; gene expression analysis; probe;

multitarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

epilepsy; cancer; ss.

W0200157275-A2.

09-AUG-2001.

30-JAN-2001; 2001wo-0800667.

04-FEB-2000; 2000us-0180312.

26-MAY-2000; 2000us-0207456.

30-JUN-2000; 2000us-0608408.

04-AUG-2000; 2000us-0632366.

21-SEP-2000; 2000us-0234687.

27-SEP-2000; 2000us-0236459.

04-OCT-2000; 2000us-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR.

WPI; 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human
brains -
Example 4, SEQ ID NO: 12792; 650bp + Sequence Listing; English.
The present invention provides a number of single exon nucleic acid
probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX
SQ Sequence 586 BP; 128 A; 165 C; 160 G; 133 T; 0 other;

Query Match 11.0%; Score 190; DB 22; Length 586;

Best Local Similarity 100.0%; Pred. No. 2,86-38;

Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 cagcttcacagacttcccccagccttgggagagatgggagagagagagagagagagag 687

DB 134 cagcttcacagacttcccccagccttgggagagatgggagagagagagagagagagag 252

QY 668 aggttgccttgcctggccagcttgcagatggagagcttccagaatgctctcatgattcc 747

DB 273 aggttgccttgcctggccagcttgcagatggagagcttccagaatgctctcatgattcc 312

QY 748 cagcttgccttgcctggccagcttgcagatggagagcttccagaatgctctcatgattcc 807

DB 313 cagcttgccttgcctggccagcttgcagatggagagcttccagaatgctctcatgattcc 372

QY 808 gcaccacagag 817

DB 373 gcaccacagag 382

Search completed: September 4, 2002, 07:43:41
Job time: 10982 sec

The first part of the paper discusses the importance of the study of the history of the United States. It is argued that the study of history is essential for a full understanding of the present. The second part of the paper discusses the importance of the study of the history of the United States. It is argued that the study of history is essential for a full understanding of the present.

QY	1401	actctcaatctctctcaaaagcaaccccaatggctgcagagcaaaacatctcaagc	1460
Dh	295	ggctctgaaattctccatgaaacacgcccacatgctctcattctgagctaaacatc	226
QY	1461	actctcaatctctctcaaaagcaaccccaatggctgcagagcaaaacatctcaagc	1520
Dh	235	tccttgcgcaaaaacacacatctccacacacacacacacacacacacacacacac	176
QY	1521	ctaaagctctcaaaagcaaccccaatggctgcagagcaaaacatctcaagc	1580
Dh	175	ttgaaattctctgaaac	116
QY	1581	actctcaatctctctcaaaagcaaccccaatggctgcagagcaaaacatctcaagc	1640
Dh	115	actctcaatctctcaaaagcaaccccaatggctgcagagcaaaacatctcaagc	56
QY	1641	ggcaaaatctctcaaaagcaaccccaatggctgcagagcaaaacatctcaagc	1698
Dh	55	ggcaaaatctctcaaaagcaaccccaatggctgcagagcaaaacatctcaagc	1

RESULT	3
A1937353	719 bp
Locus	mRNA
DEFINITION	linear EST AB-MAH-2000
wg76A08.x] NCBI GAP Project Homo sapiens cDNA clone IMAGE:2467671	

ACCESSION A1937353
VERSION A1937453 1
KEYWORDS EST
EST 5676223

REFERENCE
Baker L (to 719)
NCT01876744
TITLE
National Cancer Institute / National Institute of Neurological Disorders and Stroke. Brain Tumor Genome Anatomy Project (COGATRICAP), Tumor Genome Index
JOURNAL
Unpublished (1998)
COMMENT
Cancer Patient Registry. Ph.D.

Ph.D.
CDNA Library Preparation: M. Beatty Soares, Ph.D.; M. Farina
Ronald, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Not available; clone distribution can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.biology.wustl.edu/seq/brn/seq.htm
Insert length: 807 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 463.
Low quality sequence: 463-464

```

/oranism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2467671"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="diaplastic oligodendrocytoma"
/lab_host="DH10B"
/vector="pOrange_braint_Vect-1_pT73-RFv (pharmacia) with a
modified polylinker. Site_1, Not I; Site_2, Eco RI, 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'-
TGTCACATCTCAAGTCAGCGCAGCCGACGGGACAGGTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Pvu RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Inoue
Sources and M.Fallina-Pomato."

```

ORIGIN

Query Match	35.66;	Score 629.8;	DB 9;	Length 719;
Best Local Similarity	94.68;	Pred. No. 2.7e-87;		
Markov's 4 th Order Test	0;	Mismatches 66;	Indels 3;	Gaps 3;

[illegible]

RESULT	4
AM338865/c	
LOCUS	605 bp mRNA linear FST 31-Jan-2000
DEFINITION	hae7907_xi N ₂ -Lysan-Lent Homo sapiens cDNA clone IMAGE:287861.2 3' similar to gb:E62648 TRANCOBAMIN II PRECURSOR (HUMAN) mRNA sequence.
ACCESSION	AM338865
VERSION	AM338865.1 GI:6835491
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eulimeta; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 605)
 AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ccap>
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 JOURNAL Tumor Gene Index
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@fcrfmail.nih.gov
 Title Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 clone distribution: NCI-CCAP clone distribution information can be
 found through the L.M.A.G.E. Consortium/LNL at:
www.bio.lind.gov/bio/image/image.html
 Seq primer: -400p from cDNA
 High quality sequence stop: 415.
 Location/Qualifiers
 1..605
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGC:2878812"
 /clone_lib="NCI-CCAP_Pan1"
 /issue_type="adenocarcinoma"
 /lab_host="BH108"
 /note="Organ: pancreas; Vector: pGV-SHORT6; Site: 1; Salt: Site 2; Not: cloned unidirectionally; Primer: oligo at: Average insert size 1.72 kb; Title Technologies catalog #: 11548-013"
 BASE COUNT 126 a 164 c 190 g 120 t 5 others
 ORIGIN
 source
 1..605
 Query Match 44.8% Score 590.8; DB 9; Length 605;
 Best local Similarity 98.1%; Pred. No. 2,76-81;
 Matches 592; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 1097 aaagattgctgactacagcccaagaaatggaagaccattgagctgaagctgaatgact 1156
 605 AAGGATATGCTGACTACAGCCCAAGAAATGGAAGACCATTGAAATGAAATGGATTAGT 546
 1157 gtagccctgaagctccctcatcccaagagctggacactccctgaagcttaccctcc 1216
 545 GTAGCCCTGAAGCTCCCTCATCCCAAGAGCTGGACACTCCCTGAAGCTTACCCTCC 485
 1217 cctctatgctccctgaagaggaactgcctgagcctgctgagcctctgagcctttg 1276
 485 CTCTCTATGCTCCCTGAAGAGGAACCTGCCTGAGCCTGCTGAGCCTCTGAGCCTT 426
 1277 gataatgac 1336
 425 AGTAATGAC 495
 1337 acctggaagcgaagacgaagacatctctctggaagatctctctcagcgaagctgac 1396
 365 ACCTTGAGAGCGAAGACGAAGACATCTCTCTGGAAAGTCTTCTGAGCAAGTGGAC 306
 1397 cctgaagcctgaagctctcccatgaagcgaagcgaagctgagctgagcgaagcagatc 1456
 405 CCTGAAGCCTGAAGCTCTCCCATGAAGCGAAGCGAAGCTGAGCTGAGCGAAGCAGAT 246
 1457 aagactcttgacaaagaacgaatctgcaagcgcagagtgatggaagacgaagctctct 1516
 245 AAGACTCTTGACAAAGAACGAATCTGCAAGCGCAGAGTGATGGAAGACGAAGCTCT 186
 1517 gtaattatgaatctctgcaagagacacacacacacacacacacacacacacacac 1576
 185 GTATTATGAATCTCTGCAAGAGACACACACACACACACACACACACACACACAC 126
 1577 cctgactctgactgataatgagcagctccagctgaatggaacgaagcgaagcgaagc 1636
 125 CTCGACTCTGACTGATAATGAGCAGCTCCAGCTGAGTGTGGCAAGAGAGAGAGAG 66
 1637 ctgactctgactgataatgagcagctccagctgaatggaacgaagcgaagcgaagc 1695
 65 CTCGACTCTGACTGATAATGAGCAGCTCCAGCTGAGTGTGGCAAGAGAGAGAGAG 7

RESULT 5
 A1742486/c 649 bp mRNA linear EST 14-DEC-1999
 LOCUS w43606.x1 Scores_NSF_FR_9M_UT_PA_F_51 Homo sapiens cDNA clone
 DEFINITION IMAGE:2457874.3' similar to BC102546 (HANSBUEHLMAN 11 FEBRUARY
 (HUMAN): mRNA sequence.
 ACCESSION A1742486
 VERSION A1742486.1 GI:5110774
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 649)
 AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ccap>
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@fcrfmail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (lind.gov) for further information.
 Insert length: 730 Std Error: 0.00
 Seq primer: -400p from cDNA
 High quality sequence stop: 464.
 Location/Qualifiers
 1..649
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGC:2457874"
 /clone_lib="Scores_NSF_FR_9M_UT_PA_F_51"
 /lab_host="BH108"
 /note="Organ: pooled; Vector: p773D-pac (Pharmacia) with a modified polylinker; Site: 1; Not 1; Site 2; Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCP-enriched cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Scores NBHSF pool 1: 309384-310919, 123208-125098 Scores NB2HP pool 1: 145072-147945, 147728-148183, 149872-149255, 14902 - 150407, 151176-152257 Scores NB2HP-9W pool 1: 758280-760583, 772194-774487 Scores NBHA pool 1: 394774-407411, 420156-422875, 425890-426653 Scores NBHOT pool 1: 747276-748497, 750000-749999 Suppression by honro Scores and M. Fatima Bonaldo."
 BASE COUNT 135 a 172 c 211 g 128 t 3 others
 ORIGIN
 source
 1..649
 Query Match 34.3% Score 590.6; DB 9; Length 649;
 Best local Similarity 98.1%; Pred. No. 2,86-81;
 Matches 607; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
 1078 cccagaccgccactgctgcaagatattgactacagaccccaaggaaggagaaccatt 1137
 626 CCCAGCCGCCACTGCTTGCAAGATATGCTGCAAGATGCAAGGATGCAAGAACCAT 567
 1138 gattatgagctgagtagctgagtagcctcctggaacacacacacacacacacacac 1197
 565 GATTATGAGCTGAGTAGCTGAGTAGCCTCCTGGAACACACACACACACACACACAC 507
 1198 cctgagactctaccccccctcctgactgactcgaagacaggaagctgaagctactga 1257
 506 CCTGAGACTCTACCCCCCCTCCTGACTGACTCGAAGACAGGAAGCTGAGCTGAG 447
 1258 cctaccccccctgactctgaggaattctcctggaatcaccacagcgaagcctctgga 1317
 446 CCTACCCCCCCTGACTCTGAGGAATCTCCTGGAATCACCACAGCAGCAAGACCTTGA 387

• • •

[illegible]

RESULT	6
AA568961	
DEFINITION	AA568961 366 bp mRNA linear EST 09-SEP-1997
LOCUS	m22808.s1 NC1_CGAP_1n2 Homo sapiens cDNA clone IMAGE:1061463
DESCRIPTION	similar to contains Alu repetitive element; mRNA sequence.
ACCESSION	AA568961
VERSION	AA568961.1 01-2-142015
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

(pages 1 to 366)
NCI-CTAP <http://www.ncbi.nlm.nih.gov/ncictp>
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA library preparation: David B. Kitzman, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CCGAP clone distribution information can be
found through the L.M.A.G.E. Consortium/LINL at:
www-bio.linal.gov/4hrp/lmaga/lmaga.html
Insert length: 571 Std Error: 0.00
Seq primer: -40m13 fwd: F1 from Amersham
High quality sequence stop: 354.

SOURCE	
Accession	F07466
Organism	Homo sapiens
Cell_line	Cervix-HeLa
Cloned_from	IMAGE:1061463
Gene_name	"H1FOV1_1"
Tissue_type	Liposarcoma
Add_host	BH108
Note	Vector: pAMP10; mRNA made from liposarcoma, cDNA made by oligo-dT priming. Non-directionally cloned. Size selected on agarose gel, average insert size 600 bp.
Reference	Krizman et al. (1996) Cancer Research 56:5480-5483
BASE COUNT	84 A 98 C 72 G 112 T
ORIGIN	

Query Match 7.58% Score 914 FR 9 Length 366,
Posterior Similarity 79.68; Prob. No. 3.2e-29;
Matches 252 Conserved 0 Mismatches 0 Gaps 0
UY • YS4 LTTATTATTTTAAACCAAGGTCTACTTTTCACCAACGCTATATGATGGTAT 993

[illegible]

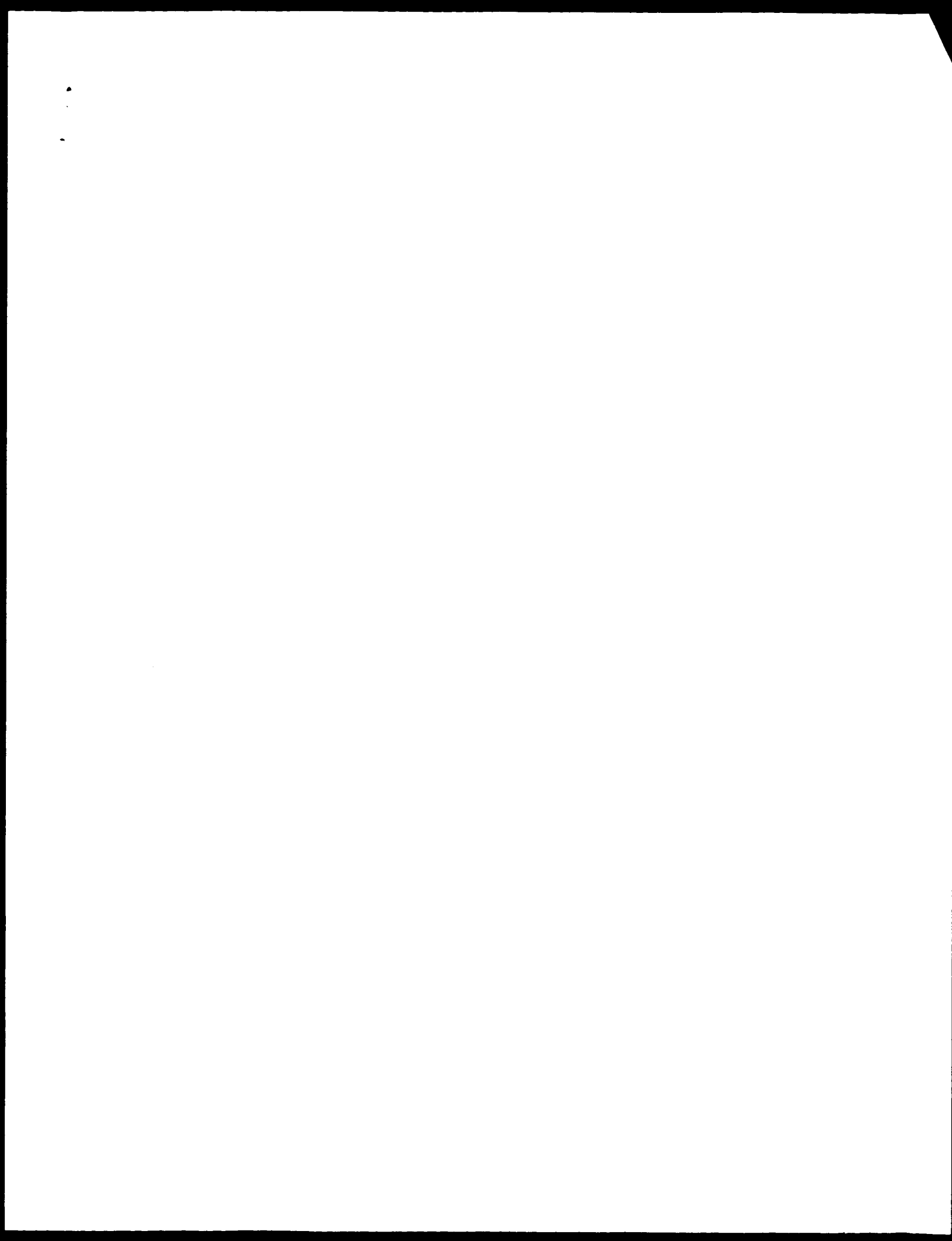
RESULT	7
AOT52148	
LOCUS	843 bp DNA Linear GSS 19-JUL-1999
DEFINITION	HU-571.B2.202.F7A BPT.L12 Human Male HMC Library Homo sapiens genomic clone Plate-1147 Col-4 low fl. DNA sequence.
ACCESSION	AOT52148
VERSION	AOT52148.1 GI:5539306
KEYWORDS	GSS
SOURCE	human
ORGANISM	Homo sapiens

REFERENCE	TITLE
MAHAIRES, G. G., WALLACE, J. C., SMITH, K., SWARTZELL, S., HOLZMAN, T., KELLER, A., SHAFER, R., FURLONG, J., YOUNG, J., ZHANG, S., ADAMS, M. D., and HODD, L.	Sequence-tagged connectors: A sequence approach to mapping and comparing the human genome

SOCIETY
 MEDLINE
 COMMENT
 11000 Mahatras Avenue, Suite 100, Seattle, WA 98104 (1997)
 99480589
 Contact: Mahatras GJ, Wallace JC, Hood L.
 High Throughput Sequencing Center
 University of Washington
 441 Green Arise Avenue North, Seattle, WA 98104, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3897

Email: j.walder@u.washington.edu
 Clones are derived from the human *CCF* library (No. 1-11). For PAC library availability, please contact plater@u.washington.edu. Clones may be purchased from SAI@u.washington.edu. For more information, visit <http://www.hsc.u.washington.edu>.
 Plate: 1147 row: H column: 4
 Seq primer: 17
 Class: BAC ends
 High quality sequence stop: 84%
 Location/Qualifiers
 1 84%

/cvecname "X6035606"
 /cfile "plac117Col 4 row 11"
 /clone.lib "R6C1-11 Human Kilo. BAC library"
 /sex "male"
 /bactec Vector: pBA7.6, Site1: EcoRI, Site2: EcoRI,
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI-Methylase. Size selected DNA was cloned into the
 pBA7.63 vector at EcoRI sites



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

COMMUNITARIANISM

Run no.: September 4, 2007, 69-49-41, South Lake Tahoe, California
 (without alignment's)
 GGG AGG MTTT CCGTTCCTT

Title:	US-09-841-158-5_COPY_1_2839
Perfect score:	2839

Sequence: TATGAGTATGGGAATTATTTCTTCACCAACTCAGTGCGAAGTAAGTGTGAAAGGAGTCCGAGTGCAGGCTG

Scoring table: IDENTITY_NUC

1. *Introduction*
 2. *Methodology*
 3. *Results*
 4. *Discussion*
 5. *Conclusion*
 6. *References*
 7. *Appendix*
 8. *Index*
 9. *Table of Contents*
 10. *Summary*
 11. *Abstract*
 12. *Keywords*
 13. *Subject Headings*
 14. *Notes*
 15. *Footnotes*
 16. *References*
 17. *Appendix*
 18. *Index*
 19. *Table of Contents*
 20. *Summary*
 21. *Abstract*
 22. *Keywords*
 23. *Subject Headings*
 24. *Notes*
 25. *Footnotes*
 26. *References*
 27. *Appendix*
 28. *Index*
 29. *Table of Contents*
 30. *Summary*
 31. *Abstract*
 32. *Keywords*
 33. *Subject Headings*
 34. *Notes*
 35. *Footnotes*
 36. *References*
 37. *Appendix*
 38. *Index*
 39. *Table of Contents*
 40. *Summary*
 41. *Abstract*
 42. *Keywords*
 43. *Subject Headings*
 44. *Notes*
 45. *Footnotes*
 46. *References*
 47. *Appendix*
 48. *Index*
 49. *Table of Contents*
 50. *Summary*
 51. *Abstract*
 52. *Keywords*
 53. *Subject Headings*
 54. *Notes*
 55. *Footnotes*
 56. *References*
 57. *Appendix*
 58. *Index*
 59. *Table of Contents*
 60. *Summary*
 61. *Abstract*
 62. *Keywords*
 63. *Subject Headings*
 64. *Notes*
 65. *Footnotes*
 66. *References*
 67. *Appendix*
 68. *Index*
 69. *Table of Contents*
 70. *Summary*
 71. *Abstract*
 72. *Keywords*
 73. *Subject Headings*
 74. *Notes*
 75. *Footnotes*
 76. *References*
 77. *Appendix*
 78. *Index*
 79. *Table of Contents*
 80. *Summary*
 81. *Abstract*
 82. *Keywords*
 83. *Subject Headings*
 84. *Notes*
 85. *Footnotes*
 86. *References*
 87. *Appendix*
 88. *Index*
 89. *Table of Contents*
 90. *Summary*
 91. *Abstract*
 92. *Keywords*
 93. *Subject Headings*
 94. *Notes*
 95. *Footnotes*
 96. *References*
 97. *Appendix*
 98. *Index*
 99. *Table of Contents*
 100. *Summary*
 101. *Abstract*
 102. *Keywords*
 103. *Subject Headings*
 104. *Notes*
 105. *Footnotes*
 106. *References*
 107. *Appendix*
 108. *Index*
 109. *Table of Contents*
 110. *Summary*
 111. *Abstract*
 112. *Keywords*
 113. *Subject Headings*
 114. *Notes*
 115. *Footnotes*
 116. *References*
 117. *Appendix*
 118. *Index*
 119. *Table of Contents*
 120. *Summary*
 121. *Abstract*
 122. *Keywords*
 123. *Subject Headings*
 124. *Notes*
 125. *Footnotes*
 126. *References*
 127. *Appendix*
 128. *Index*
 129. *Table of Contents*
 130. *Summary*
 131. *Abstract*
 132. *Keywords*
 133. *Subject Headings*
 134. *Notes*
 135. *Footnotes*
 136. *References*
 137. *Appendix*
 138. *Index*
 139. *Table of Contents*
 140. *Summary*
 141. *Abstract*
 142. *Keywords*
 143. *Subject Headings*
 144. *Notes*
 145. *Footnotes*
 146. *References*
 147. *Appendix*
 148. *Index*
 149. *Table of Contents*
 150. *Summary*
 151. *Abstract*
 152. *Keywords*
 153. *Subject Headings*
 154. *Notes*
 155. *Footnotes*
 156. *References*
 157. *Appendix*
 158. *Index*
 159. *Table of Contents*
 160. *Summary*
 161. *Abstract*
 162. *Keywords*
 163. *Subject Headings*
 164. *Notes*
 165. *Footnotes*
 166. *References*
 167. *Appendix*
 168. *Index*
 169. *Table of Contents*
 170. *Summary*
 171. *Abstract*
 172. *Keywords*
 173. *Subject Headings*
 174. *Notes*
 175. *Footnotes*
 176. *References*
 177. *Appendix*
 178. *Index*
 179. *Table of Contents*
 180. *Summary*
 181. *Abstract*
 182. *Keywords*
 183. *Subject Headings*
 184. *Notes*
 185. *Footnotes*
 186. *References*
 187. *Appendix*
 188. *Index*
 189. *Table of Contents*
 190. *Summary*
 191. *Abstract*
 192. *Keywords*
 193. *Subject Headings*
 194. *Notes*
 195. *Footnotes*
 196. *References*
 197. *Appendix*
 198. *Index*
 199. *Table of Contents*
 200. *Summary*
 201. *Abstract*
 202. *Keywords*
 203. *Subject Headings*
 204. *Notes*
 205. *Footnotes*
 206. *References*
 207. *Appendix*
 208. *Index*
 209. *Table of Contents*
 210. *Summary*
 211. *Abstract*
 212. *Keywords*
 213. *Subject Headings*
 214. *Notes*
 215. *Footnotes*
 216. *References*
 217. *Appendix*
 218. *Index*
 219. *Table of Contents*
 220. *Summary*
 221. *Abstract*
 222. *Keywords*
 223. *Subject Headings*
 224. *Notes*
 225. *Footnotes*
 226. *References*
 227. *Appendix*
 228. *Index*
 229. *Table of Contents*
 230. *Summary*
 231. *Abstract*
 232. *Keywords*
 233. *Subject Headings*
 234. *Notes*
 235. *Footnotes*
 236. *References*
 237. *Appendix*
 238. *Index*
 239. *Table of Contents*
 240. *Summary*
 241. *Abstract*
 242. *Keywords*
 243. *Subject Headings*
 244. *Notes*
 245. *Footnotes*
 246. *References*
 247. *Appendix*
 248. *Index*
 249. *Table of Contents*
 250. *Summary*
 251. *Abstract*
 252. *Keywords*
 253. *Subject Headings*
 2

Searched: 174444, 009, 00945720, 205:440

Total number of hits excluding oblique parameters 242,967

Minimum seq length: 0
Maximum seq length: 300000000

Maximum seq length: 2000000000

[illegible]

Maximum Match: 100%

listing first 45 summaries

Debt: 1000000

[illegible]

PROD. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length (bp)	to	Description
1	1794.9	63.9	8977	24	AA133787
2	1631.8	57.5	8577	24	AA133786
3	589.4	20.8	2667	92	AAH982359
4	569	20.0	569	22	AAH63875
5	646	20.0	646	22	AAV31054
6	569	20.0	649	22	AAV13394
7	569	20.0	669	22	AAK38101
8	569	20.0	569	22	AA118675
9	569	20.0	569	22	AA143998

[illegible]

[illegible][illegible]

07	241	lqatqfqaafatqaaatqaaatqfqqqgaqccaaqqaatcccaaccaaaqaaat	300
08	3247	lqatqatqqaatlaaqtadaaqqqtqgggggaqlaaqtaatlattlaagaaqaa	3306
09	401	qatccatqaaacatlaacatqatcccccgaagqccaaacqtcgaatccgaq	360
10	4307	qatlatqalatalattlaltatqatlltlltlaaqqttlaaagctggcglltggq	3356
07	361	aaatqaaqcccaatqaccccccctcccaacacccgaaggaqtcctccgaq	420
10	3367	aaatqgaatlltlatqatlltlltlltlltlaaqaatltaagggaqlltlltga	3426
09	421	lqacacatccagccttccgaatcgcaacacacatlatlaaatlttaaatltccg	480
10	3427	lqatlatcatltaqlltlltlltggatggaaatlaltaltaltlaaalttlaagtaq	3486
09	481	laaacaacacacatqatqataltlatatqtaaaqgccttqatltqtaaaatccg	540
10	3487	laaataaataacqtlqatqataltlatatqtaaaqgttggatggaaatlaag	3546
09	541	ctcaaaagatltqcttadaaaatqaaacttqgaatttaccataaaataltqcat	600
10	4547	cttadaaqaatltqtladaaqaqqaatlttqataltlatlaaataatltlatq	3606
09	601	caaatltqgcctqaaatqatqgaatalttaaaqtaatqatqtcctcaaaatlaq	660
10	4607	laqatltqatqaaatqatqgaatalttlaaqaatlataqatlttltataatlaq	3666
07	661	ctgaaggaatqatqaaatqatcttgcctaacctccatcccaagaaatlaa	720
10	4667	ctgaaggaatqatqaaatqatcttgcctaacctccatcccaagaaatlaa	3726
09	721	cttcaaatcaaaqatlaaqtacccggccaagtccaaqccacacatqaaatcc	780
10	4727	cttcaaatcaaaqatlaaqtacccggccaagtccaaqccacacatqaaatcc	3786
09	781	qatataatqatccctqatqgaatqgaatqgaatcttqatccctccat	840
10	4787	qatataatqatccctqatqgaatqgaatqgaatcttqatccctccat	3846
07	841	cttcaaaatlttlltlltlltlltlltlltlltlltlltlltlltlltlltllt	900
10	3847	cttcaaaatlttlltlltlltlltlltlltlltlltlltlltlltlltllt	3906
09	901	qatltqatccatccatccatccatccatccatccatccatccatccatccat	960
10	3907	qatltqatccatccatccatccatccatccatccatccatccatccatccat	3966
07	961	caatltqacccaaqatqaaatqgaatqatlaaqtccctcaagatqaaatcc	1020
10	3967	caatltqacccaaqatqaaatqgaatqatlaaqtccctcaagatqaaatcc	4026
09	1021	cttcaaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaa	1080
10	4027	cttcaaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaa	4086
09	1081	cccaac	1140
10	4087	cttcaaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaa	4146
07	1141	qatlttcccaatqatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaa	1200
10	4147	qatlttcccaatqatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaa	4206
09	1201	ctcaac	1260
10	4207	cttcaaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaa	4266
07	1261	cttcaaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaa	1320
10	4267	cttcaaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaaatqaa	4326

[illegible]

XX	Human sapiens.
FN	W0200157277-A2.
FD	09 AUG-2001 .
XX	
FE	30-JAN-2001 : 2001WO-US00669.
XX	
FR	04-FEB-2000 : 2000US_0180412.
FR	26 MAY-2000 : 2000US-0207456.
FR	30-JUN-2000 : 2000US-0606408.
FR	03-AUG-2000 : 2000US-0632166.
FR	21-SEP-2000 : 2000US-0234827.
FR	27-SEP-2000 : 2000US-0236459.
FR	04-OCT-2000 : 2000CH_2024263.
XX	
PA	(Male.) MOLECULAR DYNAMICS INC.
P1	Peon SG; Hancek UK; Chen W; Kark DE;
XX	
DH	WPI: 2001-48447/5Z.
XX	
FT	Human genome-derived single exon nucleic acid probes useful for
ET	analyzing gene expression in human fetal liver
XX	
PS	claim 1 : SEQ ID NO 12180; 639pp + sequence listing; English.
XX	
CC	The invention relates to a single exon nucleic acid probe for
CC	measuring human gene expression in a sample derived from human fetal
CC	liver. The single exon nucleic acid probes may be used for predicting,
CC	measuring and displaying gene expression in samples derived from human
CC	fetal liver. The present sequence is a single exon nucleic acid
CC	probe of the invention.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/databases_FWI_sequences
XX	
SC	Sequence 569 BP; 162 A; 126 C; 153 G; 128 T; 0 other;
<hr/>	
25677 Match:	25.0%; Score 579; CR 22; Length 579;
Best local Similarity:	100.0%; Pred.No: 2; 30-120;
Matches: 569;	Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;
UY	21 09 caatttttgagcttaacttggagatctgagatcctaagaacatctctgtatgttatatatctt 2258
DB	
	1 caatttgaacctagctggagacccaagatataagaacaacatattggtattgatattatattt 60
CY	22 59 taaattctcaaaaacaaaattctcaattaatctattctattctaacacattcttcgc 2318
DB	
	1 tgattctccaaacaaagaattctcgcttaattcagctattcagctattgcttcacacattatcagg 120
UY	24 19 gaaatgcttagaccttt 2378
DB	
	1 gaaatgcttagaccttt 180
CY	24 79 gaataaaattctaaaccttt 2438
DB	
	1 aaagaacattctaaaccttt 240
CY	24 99 agaaatgatttatgtctctctcaaccctaaagtctcaacatctttaaagaccttaagctggag 2498
DB	
	1 agaaatgatttatgtctctctcaaccctaaagtctcaacatctttaaagaccttaagctggag 300
CY	24 99 gaagatcgccttaagctcaacatctttaaagctggagctggagctggagctggagctggagct 2558
DB	
	1 gaagatcgccttaagctcaacatctttaaagctggagctggagctggagctggagctggagct 360
CY	25 59 ctctaaataaattaaaaaaatttadaaaatagctggagatctggagaaactcttgaagacca 2618
DB	
	1 ctctaaataaattaaaaaaatttadaaaatagctggagatctggagaaactcttgaagacca 420
CY	26 19 ataagatgaatcttgtatctttaattctaaagtgaagaaaataatgaatcttgagcaaaaaggaat 2678

D6	421	agaaggaacacttgaggcttgaacctcgaaqgaadaaqaagactcgcgcaqdaaqaatttgg	489
XX			
O7	2679	ggagacacgtctatggcgcgcaggaatgaagcttcgaagatccctgccctcgaactctg	2748
XX			
D6	481	ggacacacattcgaagcaacaqgaagtgaagcttcgaagatccctgcctgcctcgaactctg	540
OY	2749	tgccecaagagatctcttgaagcacacatc	2767
XX			
D6	541	tgccecaagagatctcttgaagcacacatc	569
RESULT	5		
AHA31054			
ID	AHA31054 standard; DNA; 569 BP.		
XX			
AC	AHA31054;		
XX			
CT	24-JAN-2002 (first entry)		
XX			
DE	Probe #5720 for gene expression analysis in human heart cell sample.		
XX			
KW	Human, gene expression; heart; microarray; vascular system; probe;		
KM	cardiovascular disease; hypertension; cardiac arrhythmia;		
RN	congenital heart disease; ss.		
XX			
OS	Homo sapiens.		
XX			
FN	W0200157274-A2.		
XX			
P3	09 AUG 2001.		
XX			
PE	30-JAN-2001; 2001MO-US00666.		
XX			
PR	04-FEB-2000; 2000MS-0180312.		
PF	24 MAY 2000; 2003MS-0207456.		
PR	30-JUN-2000; 2000MS-0608408.		
PF	03-AUG-2000; 2000MS-0632366.		
PR	21-SEP-2000; 2000MS-0234687.		
PR	27-SEP-2000; 2000MS-0236359.		
PE	04-OCT 2000; 2000SR-0024263.		
XX			
FA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX			
F1	FROM: SG, Hanzel DK, Chen W, Park DK;		
XX			
D9	WPJ; 2001-468699/53.		
PT	Single exon nucleic acid probes for analyzing gene expression in human hearts -		
Claim 1:	SEQ ID No 9520; 530PP; English.		
The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g., cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.			
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/patdb/published_patseq.htm .			
Sequence 569 BP; 162 A; 126 C; 153 G; 128 T; 0 other?			
Query Match	20.0%	Score 569; DB ZS; Length 569;	
Best Local Similarity	100.0%; E-vald. No. 2.7e-120;		
Mismatches	569; Conservative	0; Mismatches	0; Indels
		0; Gaps	0;

DE	Human bone marrow expressed single exon probe SEQ ID NO: 12658.
XX	
KM	Human bone marrow expressed exon; gene expression analysis; probe;
XX	mRNA; tissue: cartilage, adipose, lymphoma, skeletal ss.
XX	
G5	Bone sarcoma.
XX	
WM	W0209157276_A2.
FM	
XX	
PD	09 AUG 2001.
XX	
FE	40-JAN 2001; 2001-W-0500949.
XX	
FR	04-FEB 2000; 2000US-0180312.
XX	
FR	26-MAY-2000; 2000US-0207456.
FR	30-JUN-2000; 2000US-0608408.
FR	03-AUG-2000; 2000US-0632366.
FR	21-SEP-2000; 2000US-0234687.
FR	27-SEP-2000; 2000US-0736359.
FR	04-OCT-2000; 2000GB-0024263.
XX	
FA	(MFE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
WP	WO/ 2001-498900/53.
XX	
PI	Human genome derived single exon nucleic acid probes useful for
XX	analyzing gene expression in human bone marrow -
XX	
IS	Example 4: SEQ ID NO: 12658; 658bp + sequence listing; English.
XX	
CR	The present invention provides a number of single exon nucleic acid
CR	probes which are derived from genomic sequences expressed in the human
CR	bone marrow. They can be used to measure gene expression in bone marrow
CR	samples, which may enable the improved diagnosis and treatment of cancers
CR	suscept as lymphoma, leukemia and myeloma. The present sequence is set forth
XX	the probes of the invention.
SO	Sequence 569 BP; 162 A; 126 C; 154 G; 128 T; 0 other.
<hr/>	
Query Match: 20.0%; Score 569; DB 22; Length 569;	
Best Local Similarity: 100.0%; Prod No: 2 ke129;	
Matches: 569; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;	
CY	2195 cagtttgggttaagctggacctcaacatttaacacatcctatgtagtatgaatatggt 2258
DB	1 caatttgagcttagcgcccttcagcatltaaacatactctatbtgatgtatatggt 60
OY	2255 taactctcccagaacaatactcccttaactcaactcaactcaactcaactcaggc 2318
DB	61 tgactctggcggaagaagatccgtttaattgaatgtagtgcagtgcagatrragga 120
CY	2415 gcatactgaaccttctcatatcaaaacacatgltlaacctllggggaaaatgaacat 2378
DB	121 gcatactgaaccttctcttgatcaaatccacagtgtagaacctllgggaaacagat 180
OY	2475 gaacacatctaatcccttgaatctcaatgaatgaatgaatgaatgaatgaatgaat 2438
DB	181 gaacacatctaatcccttgaatctcaatgaatgaatgaatgaatgaatgaatgaat 240
CY	2495 gaatgctttatgaactccacacatctcaatgtagcttaggaacaataagaaacacct 2498
DB	241 agaatgctttatgaactccacacatctcaatgtagcttaggaacacacacacacac 300
OY	2495 gaatgctttatgaactccacacatctcaatgtagcttaggaacaataagaaacacct 2558
DB	401 gaatgctttatgaactccacacatctcaatgtagcttaggaacaataagaaacacct 360
OY	2555 ctacaaatcaaaaaaatcaaaatcagctgaatgagagagagagagagagagagagag 2618
DB	461 ctacaaatcaaaaaaatcaaaatcagctgaatgagagagagagagagagagagagag 420

```

07      2619 aggtatgaacttgagacgttaaaagtctgaaaggaaaagaagagatcgtcgtagaaagagatggg 2678
08      ||||||
09      421 gagatgaaccttgaagatctaaatcccaagataaagaaagaaatgagatgagtcagagaaagatggg 480
10      2679 gaacaacattctatgaatgatgaatgaatgagttctgaaatctctgcctgcctgaatctg 2738
11      ||||||
12      481 ggaacacattctatgaatgatgaatgaatgagttctgaaatctctgcctgcctgaatctg 540
13      541 tgcctcaaatatctatctttaaaccagctc 599
14
15 RESULT 8
16 AAT18875
17 ID AAT18875 standard; HNA; 569 BP.
18 AC AAT18875;
19 DT 12-OCT-2001 (first entry)
20 DE Probe #8608 for gene expression analysis in human cervical cell sample.
21 KW Probe; human; microarray; gene expression; cervical epithelial cell;
22 KW cervical cancer; ss.
23 OS Homo sapiens.
24 GN W0200157278-A2.
25 PN XX
26 PD 09-AUG-2001.
27 PE 30-JAN-2001; 2001W0-US00670.
28 PF XX
29 PT 04-FEB-2000; 2000US-0180412.
30 PR 26-MAY-2000; 2000US-0207456.
31 PR 30-JUN-2000; 2000US-0608408.
32 PR 03-AUG-2000; 2000US-0632466.
33 PR 21-SEP-2000; 2000US-0244687.
34 PR 27-SEP-2000; 2000US-0246459.
35 PR 04-OCT-2000; 2000GH-0024263.
36 PA (MCLD ) MCLDTHAP DYNAMICS INC.
37 PI Penn SG, Hanzel DK, Chen W, Rank DR;
38 WP1: 2001-488901/53.
39 DR Human genome-derived single exon nucleic acid probes useful for
40 analyzing gene expression in human cervical epithelial cells -
41 claim 25; SEQ ID No 8608; 487bp; English.
42
43 The present invention relates to human single-exon nucleic acid probes
44 (SENPs). The present sequence is one such probe. The SENPs are derived
45 from human HeLa cells. The SENPs can be used to produce a single exon
46 microarray, which can be used for measuring human gene expression in a
47 single derived from human cervical epithelial cells by co-silencing gene
48 expression, the probes are therefore useful in grading and/or staging
49 or diagnosis of the cervix, notably cervical cancer.
50 Note: The sequence data for this patent did not form part of the printed
51 specification, but was obtained in electronic format directly from WHO
52 at ftp.wipo.int/pub/published_pat_sequences.
53
54 Sequence 569 BP; 162 A; 126 G; 153 C; 128 T; 0 other;
55
56 Query Match 28.9%; Score 569; E# 22; Length 569;
57 Best Local Similarity 100.0%; Freq. No. 2,36,120;
58 Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
59 2109 cagtttgttatcttgagcaggcatcttctgctatctgatatctatctatctt 2708

```


REFERENCE	JOURNAL	DIRECT SUBMISSION
MO 63108, USA	Submitted (14-JAN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis,	
REFERENCE	TITLE	4 (bases 1 to 101720)
REFERENCE	AUTHORS	Waterston,R.H.
JOURNAL	DIRECT SUBMISSION	Submitted (01-JUL-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA		
REFERENCE	TITLE	5 (bases 1 to 101720)
REFERENCE	AUTHORS	Waterston,R.H.
JOURNAL	DIRECT SUBMISSION	Submitted (03-MAR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA		
REFERENCE	TITLE	6 (bases 1 to 101720)
REFERENCE	AUTHORS	Waterston,R.
JOURNAL	DIRECT SUBMISSION	Submitted (19-AUG-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
MO 63108, USA		
REFERENCE	TITLE	7 (bases 1 to 101720)
REFERENCE	AUTHORS	Waterston,R.
JOURNAL	DIRECT SUBMISSION	Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
MO 63108, USA		
REFERENCE	TITLE	8 (bases 1 to 101720)
REFERENCE	AUTHORS	Waterston,R.H.
JOURNAL	DIRECT SUBMISSION	Submitted (03-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA		
COMMENT		On Jul 1, 1999 this sequence version replaced gi:4156152. Center project name: H_0j56j10. Location/Qualifiers 1..101720 /orgnism="Homo sapiens" /db_xref="taxon:9606"
FEATURES	SOURCE	/contig="wt1-56j10"
BASE COUNT	ORIGIN	24683 a 26390 c 26554 g 24093 t

Query Match	99.48%	Score 2826.47	DP 97	Length 101729
Host Local Similarity	99.88%	Prod. No. 97		
Matches 2834	Conservative	0	Mismatches 4	Indels 1
				Gaps 1
QY 1	atataatagaataatgatcgtcttcctactatccctacccaccctctagcaatgacg	60		
DB 54583	ATATGTTGGGAATATGCTGCTCTCTCAATCCAACTCCCGACCCCTACGACATGAGTC	54562		
QY 91	caatgaatgaagaagagagatctctccctccttaacttgaacacctaaactctggag	120		
DB 54643	CACTTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT	54702		
QY 131	atcgaagatgaactctctgagagagatctctgctctctcctcctcctcctcctcctcct	180		
DB 54704	ATTTGAAGTCACTCTTGGAAAGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	54762		
QY 181	ggctatactcctcatggccactcctcccaagcctggcctcctcctgaaatagaatcataagt	240		
DB 54703	GATTAATCTTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACTAACT	54822		
QY 241	tgttactggaagaacagagagaagctctggagagagacaaagacatccaccacgaacaaagt	300		
DB 54833	TGTTACTGTAATAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT	54892		
QY 301	ggctcccatctgagagacatttaacacatgactccctctgaaatgctcaaaagcttgctctggg	360		
DB 54883	GGCTCCATCTGCACACACTTCATATTACTIONCTTTAAAGTTCGAACCTTGGCTTCTGGG	54942		
QY 361	aaatttgagctcccaactggcctcctcctcctctctagaacctccaaggatgctcctccctg	420		

[illegible]

420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

4131
4132
4133
4134
4135
4136
4137
4138
4139
4140
4141
4142
4143
4144
4145
4146
4147
4148
4149
4150
4151
4152
4153
4154
4155
4156
4157
4158
4159
4160
4161
4162
4163
4164
4165
4166
4167
4168
4169
4170
4171
4172
4173
4174
4175
4176
4177
4178
4179
4180
4181
4182
4183
4184
4185
4186
4187
4188
4189
4190
4191
4192
4193
4194
4195
4196
4197
4198
4199
4200
4201
4202
4203
4204
4205
4206
4207
4208
4209
4210
4211
4212
4213
4214
4215
4216
4217
4218
4219
4220
4221
4222
4223
4224
4225
4226
4227
4228
4229
4230
4231
4232
4233
4234
4235
4236
4237
4238
4239
4240
4241
4242
4243
4244
4245
4246
4247
4248
4249
4250
4251
4252
4253
4254
4255
4256
4257
4258
4259
4260
4261
4262
4263
4264
4265
4266
4267
4268
4269
4270
4271
4272
4273
4274
4275
4276
4277
4278
4279
4280
4281
4282
4283
4284
4285
4286
4287
4288
4289
4290
4291
4292
4293
4294
4295
4296
4297
4298
4299
4300
4301
4302
4303
4304
4305
4306
4307
4308
4309
4310
4311
4312
4313
4314
4315
4316
4317
4318
4319
4320
4321
4322
4323
4324
4325
4326
4327
4328
4329
4330
4331
4332
4333
4334
4335
4336
4337
4338
4339
4340
4341
4342
4343
4344
4345
4346
4347
4348
4349
4350
4351
4352
4353
4354
4355
4356
4357
4358
4359
4360
4361
4362
4363
4364
4365
4366
4367
4368
4369
4370
4371
4372
4373
4374
4375
4376
4377
4378
4379
4380
4381
4382
4383
4384
4385
4386
4387
4388
4389
4390
4391
4392
4393
4394
4395
4396
4397
4398
4399
4400
4401
4402
4403
4404
4405
4406
4407
4408
4409
4410
4411
4412
4413
4414
4415
4416
4417
4418
4419
4420
4421
4422
4423
4424
4425
4426
4427
4428
4429
4430
4431
4432
4433
4434
4435
4436
4437
4438
4439
4440
4441
4442
4443
4444
4445
4446
4447
4448
4449
4450
4451
4452
4453
4454
4455
4456
4457
4458
4459
4460
4461
4462
4463
4464
4465
4466
4467
4468
4469
4470
4471
4472
4473
4474
4475
4476
4477
4478
4479
4480
4481
4482
4483
4484
4485
4486
4487
4488
4489
4490
4491
4492
4493
4494
4495
4496
4497
4498
4499
4500
4501
4502
4503
4504
4505
4506
4507
4508
4509
4510
4511
4512
4513
4514
4515
4516
4517
4518
4519
4520
4521
4522
4523
4524
4525
4526
4527
4528
4529
4530
4531
4532
4533
4534
4535
4536
4537
4538
4539
4540
4541
4542
4543
4544
4545
4546
4547
4548
4549
4550
4551
4552
4553
4554
4555
4556
4557
4558
4559
4560
4561
4562
4563
4564
4565
4566
4567
4568
4569
4570
4571
4572
4573
4574
4575
4576
4577
4578
4579
4580
4581
4582
4583
4584
4585
4586
4587
4588
4589
4590
4591
4592
4593
4594
4595
4596
4597
4598
4599
4600
4601
4602
4603
4604
4605
4606
4607
4608
4609
4610
4611
4612
4613
4614
4615
4616
4617
4618
4619
4620
4621
4622
4623
4624
4625
4626
4627
4628
4629
4630
4631
4632
4633
4634
4635
4636
4637
4638
4639
4640
4641
4642
4643
4644
4645
4646
4647
4648
4649
4650
4651
4652
4653
4654
4655
4656
4657
4658
4659
4660
4661
4662
4663
4664
4665
4666
4667
4668
4669
4670
4671
4672
4673
4674
4675
4676
4677
4678
4679
4680
4681
4682
4683
4684
4685
4686
4687
4688
4689
4690
4691
4692
4693
4694
4695
4696
4697
4698
4699
4700
4701
4702
4703
4704
4705
4706
4707
4708
4709
4710
4711
4712
4713
4714
4715
4716
4717
4718
4719
4720
4721
4722
4723
4724
4725
4726
4727
4728
4729
4730
4731
4732
4733
4734
4735
4736
4737
4738
4739
4740
4741
4742
4743
4744
4745
4746
4747
4748
4749
4750
4751
4752
4753
4754
4755
4756
4757
4758
4759
4760
4761
4762
4763
4764
4765
4766
4767
4768
4769
4770
4771
4772
4773
4774
4775
4776
4777
4778
4779
4780
4781
4782
4783
4784
4785
4786
4787
4788
4789
4790
4791
4792
4793
4794
4795
4796
4797
4798
4799
4800
4801
4802
4803
4804
4805
4806
4807
4808
4809
4810
4811
4812
4813
4814
4815
4816
4817
4818
4819
4820
4821
4822
4823
4824
4825
4826
4827
4828
4829
4830
4831
4832
4833
4834
4835
4836
4837
4838
4839
4840
4841
4842
4843
4844
4845
4846
4847
4848
4849
4850
4851
4852
4853
4854
4855
4856
4857
4858
4859
4860
4861
4862
4863
4864
4865
4866
4867
4868
4869
4870
4871
4872
4873
4874
4875
4876
4877
4878
4879
4880
4881
4882
4883
4884
4885
4886
4887
4888
4889
4890
4891
4892
4893
4894
4895
4896
4897
4898
4899
4900
4901
4902
4903
4904
4905
4906
4907
4908
4909
4910
4911
4912
4913
4914
4915
4916
4917
4918
4919
4920
4921
4922
4923
4924
4925
4926
4927
4928
4929
4930
4931
4932
4933
4934
4935
4936
4937
4938
4939
4940
4941
4942
4943
4944
4945
4946
4947
4948
4949
4950
4951
4952
4953
4954
4955
4956
4957
4958
4959
4960
4961
4962
4963
4964
4965
4966
4967
4968
4969
4970
4971
4972
4973
4974
4975
4976
4977
4978
4979
4980
4981
4982
4983
4984
4985
4986
4987
4988
4989
4990
4991
4992
4993
4994
4995
4996
4997
4998
4999
5000

SCIENCE
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
1 (bases 1 to 191887)
Britten, R., Linton, L., Nussbaum, C., and Landier, E.
Homo sapiens chromosome 17, clone CH-2545H
Unpublished
2 (bases 1 to 191887)
Britten, R., Linton, L., Nussbaum, C., Landier, E., All, A., Allen, N.,
Anderson, S., Batra, N., Bastien, V., Beauslaux, L., Berkhalter, B.,
Brown, A., Camarero, J., Campiano, A., Chang, J., Chazotte, R.,
Chen, Y., Chelapala, M., Collins, S., Collinge, A., Cook, A.,
Cooke, P., Doolittle, R., Dwyer, S., Eickholt, S., Fager, S., Fano, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Girio, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, R., Heald, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kawai, A., Karatas, A., Kelis, C., Laporte, K.,
Lamarez, R., Landier, E., Lebeck, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McKean, P., McKernan, K., McPherson, R., Melton, J.,
Meneilly, L., Milova, T., Mienda, V., Murphy, T., Naylor, J., Nguyen, C.,
North, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Olivier, J., Peterson, K., Phukphang, P., Pierre, N., Pollard, V.,
Raymond, C., Rella, R., Rickard, M., Riley, R., Riser, C., Rongey, P.,
Roman, J., Roselli, M., Roy, A., Santos, R., Schaefer, S., Schpack, R.,
Seaman, S., Severe, P., Spencer, H., Stapp, R., Stapp, N.,
Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
Topfman, K., Travers, M., Travis, N., Trifillo, J., Vassiliev, R.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, C.,
Zacharia, J., Zambek, L., Zimmer, A., and Zody, M.
Direct Submission
Submitted (JAN-V-2001) Wellcome Trust/Wellcome Center for Genome
Research, 120 Charles Street, Cambridge, MA 02141, USA
All reports were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://trf.bioc.ox.ac.uk/cgi-bin/trf-simulated.htm
Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WILR
Web site: http://www.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project information
Center project name: L10708
Center clone name: 2545.H.1
Summary Statistics
Sequencing vector: pUC19, n/7, 100% of reads
Assembly program: Phrap, version 9.60731
Consensus quality: 182941 bases at least 40
Consensus quality: 187111 bases at least 50
Consensus quality: 188492 bases at least 60
Insert size: 195009; average=19
Insert size: 189187; sum-of-counts
Quality coverage: 11.7 in 920 bases; sum of counts
Quality coverage: 12.9 in 920 bases; sum of counts
NOTE: This is a working draft sequence. It currently
consists of 96 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
643 742: gap of 100 bp
743 1333: contig of 591 bp in length
1334 1443: gap of 100 bp
1444 2386: contig of 903 bp in length
2387 2445: gap of 100 bp
2446 3175: contig of 729 bp in length
3176 3275: gap of 100 bp

FEATURES
Source
1..191887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CH-2545H"
/clone_11b="CH19 Human BAC"
1..642
/note="assembly_fragment"
743..1333
/note="assembly_fragment"
1334..2386
/note="assembly_fragment"
2387..2445
/note="assembly_fragment"
2446..3175
/note="assembly_fragment"
3176..3275
/note="assembly_fragment"
3276..4422
/note="assembly_fragment"
4423..5142
/note="assembly_fragment"
5143..5908
/note="assembly_fragment"
5909..6008
/note="assembly_fragment"
6009..6926
/note="assembly_fragment"
6927..7026
/note="assembly_fragment"
7027..7674
/note="assembly_fragment"
7675..8461
/note="assembly_fragment"
8462..9861
/note="assembly_fragment"
9862..9186
/note="assembly_fragment"
9187..10016
/note="assembly_fragment"
10017..10116
/note="assembly_fragment"
10117..10729
/note="assembly_fragment"
10730..10929
/note="assembly_fragment"
10930..11973
/note="assembly_fragment"
11974..12073
/note="assembly_fragment"
12074..14092
/note="assembly_fragment"
14093..13192
/note="assembly_fragment"
13193..14041
/note="assembly_fragment"
14042..14131
/note="assembly_fragment"
14132..14841
/note="assembly_fragment"
14842..14931
/note="assembly_fragment"
14932..16764
/note="assembly_fragment"
16765..16694
/note="assembly_fragment"
16695..17768
/note="assembly_fragment"
17769..17688
/note="assembly_fragment"
17689..18647
/note="assembly_fragment"
18648..18737
/note="assembly_fragment"
18738..20046
/note="assembly_fragment"
20047..20136
/note="assembly_fragment"
20137..22572
/note="assembly_fragment"
22573..23723
/note="assembly_fragment"
23724..24711
/note="assembly_fragment"
24712..23931
/note="assembly_fragment"
23932..25353
/note="assembly_fragment"
25354..25453
/note="assembly_fragment"
25454..34669
/note="assembly_fragment"
34670..34769
/note="assembly_fragment"
34770..79954
/note="assembly_fragment"
79955..80054
/note="assembly_fragment"
80055..139801
/note="assembly_fragment"
139802..139901
/note="assembly_fragment"
139902..191887
/note="assembly_fragment"

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapped clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL: EMBL; SWISSPROT: Trl; TrEMBL: WPI; WIPWEP: Information on the WIPWEP database can be found at <http://www.sanger.ac.uk/Projects/US/sequences/wipwep/>. This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/Projects/US/sequences/wipwep/>.

RPI1-54923 is from the library RPI1-11.2 constructed at the Roswell Park Cancer Institute by the group of Peter de Jong. For further details see <http://www.sanger.ac.uk/Projects/US/sequences/wipwep/>.

IMPORTANT: This sequence is not the entire insert of clone RPI1-54923. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RPI1-54923 is at 18547 in this sequence. The true left end of clone RPI1-44805 is at 86817 in this sequence. The true right end of clone RPI1-32572 is at 100 in this sequence.

FEATURES

```

SOURCE
    1..185467
        /location="nullifiers"
        /format="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="6"
        /clone="RPI1-54923"
        /clone_lib="RPI1-11.2"
        1..203
repeat_region
    /note="AluB repeat: matches 97..203 of consensus"
    376..852
misc_feature
    /note="match: GSS: Em:AO7867967"
    989..1276
repeat_region
    /note="AluB repeat: matches 1..287 of consensus"
    2200..4475
repeat_region
    /note="AluB repeat: matches 3473..4141 of consensus"
    4973..5266
repeat_region
    /note="AluB repeat: matches 1..294 of consensus"
    5431..5807
misc_feature
    /note="MSTR repeat: matches 1..390 of consensus"
    5907..6258
misc_feature
    /note="match: GSS: Em:AO788293"
    6548..6686
repeat_region
    /note="AluB repeat: matches 5236..5794 of consensus"
    6735..7223
repeat_region
    /note="AluB repeat: matches 1472..2061 of consensus"
    7284..7580
repeat_region
    /note="AluB repeat: matches 1..298 of consensus"
    7581..7946
repeat_region
    /note="AluB repeat: matches 1866..1462 of consensus"
    8018..8159
repeat_region
    /note="AluA repeat: matches 2..141 of consensus"
    8303..8588
repeat_region
    /note="L1M6 repeat: matches 449..759 of consensus"
    8712..8928
repeat_region
    /note="AluB repeat: matches 85..300 of consensus"
    complement(9457..9638)
misc_feature
    /note="match: GSS: Em:AO353367"
    9668..9967
repeat_region
    /note="AluB repeat: matches 6..307 of consensus"
    10859..11098
repeat_region
    /note="L1M7 repeat: matches 5855..6060 of consensus"

```

```

repeat_region
    11099..11394
    /note="AluB repeat: matches 1..294 of consensus"
    13689..13998
    /note="AluB repeat: matches 1..311 of consensus"
repeat_region
    15065..15567
    /note="L2 repeat: matches 1818..2355 of consensus"
repeat_region
    15767..16026
    /note="AluB repeat: matches 1..289 of consensus"
misc_feature
    complement(15863..16438)
    /note="match: GSS: Em:AO517248"
    16029..16086
misc_feature
    /note="29 copies 2 met qa 698 conserved"
    complement(17204..17559)
misc_feature
    /note="match: GSS: Em:AO072859"
    17630..17991
    /note="match: GSS: Em:AO777121"
    18477..18949
    /note="AluB repeat: matches 1..304 of consensus"
    complement(19574..20078)
    /note="match: GSS: Em:AO777121"
    20331..20800
repeat_region
    /note="AluB repeat: matches 1..287 of consensus"
    21264..21560
repeat_region
    /note="AluB repeat: matches 1..295 of consensus"
    21931..22230
repeat_region
    /note="AluB repeat: matches 1..295 of consensus"
    22577..22856
repeat_region
    /note="MIR repeat: matches 155..260 of consensus"
    22821..23119
repeat_region
    /note="AluB repeat: matches 2..292 of consensus"
    23780..23903
repeat_region
    /note="L1M1 repeat: matches 5712..5886 of consensus"
    24212..24510
repeat_region
    /note="AluB repeat: matches 1..297 of consensus"
    25084..25216
repeat_region
    /note="L2 repeat: matches 3779..2710 of consensus"
    25749..25903
repeat_region
    /note="PRAM repeat: matches 9..163 of consensus"
    complement(26457..27119)
    /note="match: STS: Em:AF191939"
    27133..27822
misc_feature
    /note="match: STS: Em:AF191939"
    27339..27648
repeat_region
    /note="AluB repeat: matches 1..312 of consensus"
    28181..28418
repeat_region
    /note="AluB repeat: matches 1..298 of consensus"
    28719..28928
repeat_region
    /note="THE1B repeat: matches 1..364 of consensus"
    30450..30741
repeat_region
    /note="L1M5 repeat: matches 5574..5865 of consensus"
    31560..31703
repeat_region
    /note="L1M3A repeat: matches 5797..5932 of consensus"
    31704..32012
repeat_region
    /note="AluB repeat: matches 1..312 of consensus"
    32033..32346
repeat_region
    /note="L1M3A repeat: matches 7428..5797 of consensus"
    32831..33017
repeat_region
    /note="AluB repeat: matches 1..250 of consensus"
    33018..33048
repeat_region
    /note="AluB repeat: matches 218..248 of consensus"
    34161..34432
repeat_region
    /note="AluB repeat: matches 25..304 of consensus"
    34468..34514
repeat_region
    /note="MIR repeat: matches 196..247 of consensus"
    34854..35144
repeat_region
    /note="AluB repeat: matches 6..296 of consensus"
    35394..35445
repeat_region
    /note="MIR repeat: matches 161..250 of consensus"
    35446..35489
repeat_region
    /note="L2 repeat: matches 2633..2705 of consensus"
    36738..37033
repeat_region
    /note="AluB repeat: matches 1..295 of consensus"
    37167..37248

```